The Role of Probability in Indicative and Deontic Conditionals

Edward John Sutherland BA (Hons)

A thesis submitted in partial fulfilment of the requirements of the University of Wolverhampton for the degree of Doctor of Philosophy

April 1999

This work or any part thereof has not previously been presented in any form to the University or to any other body whether for the purposes of assessment, publication or for any other purpose (unless previously indicated). Save for any express acknowledgments, references and / or bibliographies cited in the work, I confirm that the intellectual content of the work is the result of my own efforts and of no other person.

Signature..E. J. Sutherland....

Date....29th February 2000.....

ACKNOWLEDGEMENTS

I am indebted to Professor Ken Manktelow for his expert guidance, supervision and patience throughout the completion of this thesis. His tolerance during the writing-up period should also be acknowledged here. His support throughout the running of the experiments and the production of the thesis were above and beyond what any could reasonably expect. I am also grateful to Professor David Over as second supervisor. His encouragement and assistance were often invaluable. I also acknowledge the University of Wolverhampton in awarding a grant to Professor Manktelow, which funded the research programme.

Thanks must also be accorded to a number of others. The psychology technicians at the University of Wolverhampton for their assistance with the computerised experiments reported in Chapter Seven. Special thanks must go to Yvette Easthope for her help in collecting the data for experiment ten. I must thank all my colleagues at the University of Derby, but particularly the cognitive team, for all their encouragement, support, and for the times they were prepared to listen to my latest idea.

Finally, I must thank Kirsty for her unswerving support throughout many years, for this thesis and for everything else. It would never have happened without you.

This thesis is dedicated to the memory of Rose Alice Sutherland.

6 March 1915 - 2 January 1993

CONTENTS

APPENDICES

ABSTRACT

The Role of Probability in Indicative and Deontic Conditionals

Edward J. Sutherland

Research by Manktelow and Over (1991,1992) suggested that utility, a construct usually associated with the separate discipline of decision-making research, played a role in mediating deontic reasoning. This research added to a growing feeling of links between decision-making and reasoning. In order to further these links an additional construct from decision-making research was applied to reasoning: probability.

Initial experiments considered the role of probability in indicative conditionals, and found evidence to support the use of probability by participants on tasks such as this. The task used was Wason's (1966, 1968) four-card selection task, although this task was revised in order to facilitate the use of probabilistic information in the task. This resulted in the introduction of the Large Array Selection Task (LAST).

Following these initial findings a move was made to the realm of deontic reasoning. Deontic reasoning is a form of practical reasoning about actions, reasoning about what actions one should, ought or may perform. The task used here was a revised version of Cheng and Holyoak's (1985) immigration task employing the LAST. These results showed a large effect of probability on deontic reasoning. Finally a set of computer experiments were run which presented participants with probabilistic information, and demonstrated that participants could extract probabilistic information from the data presented to them.

The interpretation of these results considers current theories of both indicative and deontic reasoning, and the mental models approach of Johnson-Laird and Byrne (1991) is favoured here. This theory can account for the findings presented whereas alternative theories made predictions that were not supported by the data presented in this thesis. Finally there is a discussion of future research including a need to research the role of probability in other related tasks, such as deontic statements involving threats and promises, as well as looking at utility and probability in inductive reasoning.

CHAPTER 1

Page

1.1 Thinking and Reasoning

1.1.1 Human Thought - Its Nature

It has been claimed that the one thing that distinguishes humans from the rest of the animal kingdom is the fact that we are rational beings (Garnham & Oakhill, 1994). We are constantly engaged in thought processes of some kind, and these can vary tremendously in their nature, structure and purpose. The opening section of this thesis will attempt to consider the various ways in which we use our thought processes, and how we apply different modes of thought to different goals in our thinking.

Johnson-Laird (1988, pp. 220) attempted to present 'a taxonomy of thought'. This structure moves through various ways of thinking including daydreaming, calculation, creation and two forms of reasoning: inductive and deductive. These two forms of reasoning are of central importance to our existence as humans. As Johnson-Laird points out they can lead to an increase in semantic information (only in the case of inductive reasoning), they are involved in our goal of increasing understanding, adding to our existing knowledge base, in order to gain a full comprehension of our environment. Our thinking impinges on every action in which we are involved: problem solving, reasoning, creativity, and daydreaming. It is therefore clear that an understanding of thinking generally, and reasoning specifically is central to a full understanding of human psychology.

Thus, it is of no surprise that psychologists have concerned themselves with understanding human thinking. The study of human thought processes dates back over 2,000 years to Aristotle (Audi, 1989) and is still of interest to many modern cognitive psychologists,

cognitive scientists and philosophers. The purpose of the introductory chapter of this thesis is to address the history of research into human thinking, although special consideration will be placed on deductive reasoning, and within this field conditional reasoning as it is the form of reasoning to be considered in the present research programme.

1.1.2 Historical Perspectives on Reasoning Research

In order for the current research to be understood fully, it is imperative that a full and clear outline of the history of reasoning research be given. It is in the opening chapters of this thesis that this shall be done. This review will begin with the importance of reasoning to human life and the history of reasoning research, followed by the patterns of performance observed, and finally consider some of the theories that have been offered to account for these patterns.

It is clear that life as we know it now would be as nothing without the ability to reason clearly about the world around us. However, the majority of reasoning that we engage in everyday seems to occur unconsciously, and we draw our inferences without knowingly being occupied in reasoning. For example, if we go to visit our parents, and the car is not in the drive, then we conclude that they are not at home, given that we know that the car is not at the garage etc. This is not a deduction that we really think about making, or have to work at to draw. We apply our world knowledge to the situation and make the above deduction, even though if one actually considers the rules that have been applied, and the reasoning that is involved, then one realises how amazing the abilities of human reasoning are.

Without the ability to draw inferences such as this, everyday conversation would be almost impossible (Manktelow and Over, 1990). For someone to tell a story or give instructions requires the listener to reason clearly. Without this ability such everyday activities would be almost impossible. Not only must the listener make inferences from the information that they are given, but the person who is giving the instructions is also aware of the inferences that the other person is able to draw, and will adjust the information that they give. For example, when speaking to a child then one does not expect the listener to draw many inferences. Consequently one tends to give the child all the necessary information to carry out the required task, whereas when speaking to an adult who knows the task well only the barest information need be given to the listener.

Therefore, it is clear that reasoning plays a huge role in the lives of everybody, the way we speak, the way we act, and the way that other people react to us. However, it is important to note here that there are different kinds of reasoning, as mentioned earlier: deductive reasoning, and inductive reasoning. Deductive reasoning is the one that concerns this thesis, and inductive reasoning will be dealt with only briefly. Deductive reasoning is a form of reasoning that does not add any further information to that which has already been given. A deductive inference is one that draws out a conclusion that is latent, or implicit in the information given. The conclusion that is drawn is not a trivial one, but one that has not been explicitly stated already, but which assumes nothing outside of the information given in the premises.

Therefore, it is important now to look at the research that has been carried out by cognitive psychologists into the area of human deductive reasoning. Then to consider some of the

theories which have been offered as an explanation for the phenomena which have been found, and also some of the shortcomings of these theories.

1.2 Reasoning Research

1.2.1. Introduction to research

As was discussed briefly above, there are various forms of thinking and reasoning, and these all serve a different purpose for humans, and many of these thought processes have been investigated in the psychology laboratory. The purpose of this section is to present a brief overview of the research into thinking and reasoning, considering the various modes of thought and some of the methods employed to study them.

1.2.2. Inductive Reasoning

Inductive inferences do not necessarily follow from the information with which we have been presented, or more specifically, they are *non-logical inferences* (Stevenson, 1993). Induction plays a major role in everyday life; induction is a form of thought that has a precise goal (as opposed to creation), and yet unlike deduction, induction leads to an increase in semantic information: that is some information that was not present in the premises. There is a lot of work involved in induction, although it seems that people are remarkably unaware that they are involved in such a complex mental process (Johnson-Laird, 1988). However, it is an essential part of human thinking, as we do not often have all the information that we need to draw the conclusions that we would like to. Consequently

we must assume some information, and this is where the role of probability comes into the equation, and it is this that is the major concern of this thesis.

For example, making an inference from a generalisation to a particular case, all swans are white, therefore this swan is white, is a deductive inference. However, the reverse process of this, which is empirical generalisation, is a classic case of everyday induction; thus we would reason inductively thus, "All the cats I have seen have tails, therefore, this cat has a tail" (see Stevenson, 1993). Generalisations such as this are, logically speaking, unsound, because the person drawing this conclusion will not have seen all the cats in the world. The classic example being that they may never have seen a Manx cat, which does not have a tail, and so the conclusion is not based on a full set of information. However, inductions are still an essential part of our thinking in the everyday world. Inductive inferences are informative to humans, although they are not guaranteed to be true. The fact that inductions are not guaranteed to be true is part of their intrinsic nature. They are *plausible* given our knowledge of the world, we bring some of our beliefs to bear on the premises with which we have been presented in order to make an inference which we believe is likely given what we know about the world.

1.2.3. Deductive Reasoning

Deduction is distinct from induction in that deductions are logically necessary. This means that a deductive inference follows from the information in the premises. For example:

All Men are mortal

Socrates was a man

∴ Socrates was mortal

The example given above is known as a syllogism, and is one of the most studied forms of deductive reasoning performance in humans (see Garnham and Oakhill, 1994). Syllogistic reasoning will be considered briefly below. The important point to note from this example is that the information about Socrates and mortality has been inferred from the information in the premises. Thus deductive reasoning is based around forms of *logic*. Logic represents a competence theory, that is how people *would* think if they never made mistakes, it is not a performance theory of how people *really* perform when engaging in deductive thought processes (Garnham and Oakhill). A central concept for consideration throughout this thesis then is logic and its relationship to human deduction.

It is important to note here that logic is concerned with the validity of conclusions, not whether they are true or not. A conclusion can be true and not valid, not true and valid, true and valid, and finally untrue and invalid. When psychologists study human reasoning they are concerned with whether participants conform to the dictates of formal logic or not. For example, participants are asked what conclusions follow (generation task) or whether a given conclusion follows (evaluation task). The responses given by participants

are then compared to the valid conclusions as dictated by logic, and the participant is judged correct or not accordingly (see Evans, Newstead and Byrne, 1993; Manktelow, 1999, for a detailed discussion of syllogistic reasoning).

This section has dealt briefly with the nature of deduction in general terms, and used syllogistic reasoning as an example. As will be discussed below humans display a wide variety of performances on deductive reasoning tasks, and numerous factors can have an effect on their reasoning performance. Attention will now be turned to the specific area of conditional reasoning, as this is the form of deduction with which this thesis is concerned.

1.2.4 Conditional Reasoning.

It is important that attention is turned to studies of conditional reasoning, in order to gain some insight into the level of reasoning competence that exists in humans. In order to do this a review of research literature displaying typical patterns of reasoning behaviour will be presented. In order to place this research in context in is necessary to consider the nature of conditional reasoning itself, what it is, what the associated inferences are, and finally to examine some of the findings of research on conditional reasoning.

Conditional reasoning is reasoning concerned with the logical connective 'if...then' (see Howson, 1997 for a brief discussion of the conditional in logic). Initially conditionals will be considered in their abstract form, to illustrate the inferences associated with them. For example:

If p then q	The conditional statement.
p	Additional information regarding one of the components of the conditional.
∴	An inference can then be drawn.

With conditionals there are four classical inferences to consider: two of which are valid, and two of which are invalid. The valid inferences shall be considered first. Given the following premises:

If p then q

p

It is possible to derive q. This rule is called modus ponens (MP), this is an easy inference to draw, as can be seen from the regularity with which both adults and children draw or endorse the inference (Wildman and Fletcher, 1977). The second valid inference that can be drawn with conditionals is the modus tollens (MT) inference. Given the premises:

If p then q

not-q

It is possible to derive not-p. This is a harder inference to draw, and is drawn or endorsed less frequently than the MP inference above (Wildman and Fletcher, 1977). As mentioned earlier there are two invalid inferences that can be drawn with conditionals, namely affirmation of the consequent (AC) and denial of the antecedent (DA). Affirmation of the consequent is as follows:

If p then q

q

∴p

This is an invalid argument form, however, it is commonly endorsed by participants in laboratory studies (Evans, 1977). The final inference to be examined here is denial of the antecedent:

If p then q

Not-p

∴Not-q

As with the AC inference drawing or endorsement rates of this inference is reasonably high, despite its being invalid. The validity of inferences can be assessed via logical truth tables. The truth table for conditional statements is given below:

Table 1.1 Truth table for conditional statement if p then q

p	q	If p then q
T	T	T
T	F	F
F	T	T
F	F	T

T = True F = False

Table 1.1 displays the truth-values for the conditional statement given the truth-values of the antecedent and the consequent. It can be seen that in all cases except where p is true and q is false the conditional is true. However, the truth table also allows assessment of the validity of argument forms, as outlined above. For the MP inference in cases where If p then q is true, and p is true (line 1 in table 1.1) it is possible to see that q is true, there is no ambiguity, and thus the argument form is valid. For the MT inference it can be seen that where the conditional is true and q is false, not-q, (line 4 in table 1.1) then p is false, with no ambiguity. However, if the AC and DA inferences are assessed via truth tables then it is clear why they are invalid. Firstly AC: Where the conditional is true, and q is also true (lines 1 and 3 in table 1.1.), here p could be true (line 1) or could be false (line 3), thus the argument form could lead to ambiguity, and thus is invalid. A similar explanation can be

applied to the DA inference. When the conditional is true, and p is false, not-p, (lines 3 and 4 from table 1.1) q could be either true or false, and thus the argument form is invalid. Given that it is clear which of these inferences are valid, and which of them are invalid, it is of interest to see how often participants in conditional reasoning tasks draw or endorse valid and invalid conclusions, through laboratory experiments. However, it is important to note that conditional statements can be interpreted in two ways - as the conditional described above, or as a biconditional, which is if and only if p then q (this is denoted in formal logic as *iff*, Howson, 1997). Given this interpretation then all four of the inferences described above are valid as the biconditional has a different truth table to that of the conditional, but biconditionals are not of interest here.

Attention will now be turned to empirical studies of conditional reasoning, and how frequently participants draw or endorse the four conclusions described above. Firstly, Rumain, Connell and Braine (1983): in this study, the researchers considered the performance of many different age groups, in order to examine the development of conditional reasoning.

Although Rumain et al. varied the nature of the problems which they presented to their participants they are presented here as averaged across these problem types in order to present a more compact set of results here.

For adults, the percentages of correct responses, pooled across problem types, were as follows: Modus ponens: 92%, denial of the antecedent: 74%, affirmation of the consequent: 69%, and finally modus tollens: 69%. For children, the percentages of correct responses, pooled across problem types, were as follows: Modus ponens: 89%, denial of the

antecedent: 40%, affirmation of the consequent: 32%, and finally modus tollens: 62%. Although results for both adults and children have been given, it is really with the adult responses that the review here is mainly concerned. As suggested briefly above there was little difficulty with the MP inference (92%), whereas the MT (69%) inference was drawn less frequently, however, this requires the application of more than one inference rule, unlike MP. However, the participants also displayed a tendency to endorse or draw the fallacious inferences (DA and AC). Therefore, although these results show far from perfect reasoning performance, they do suggest a reasonably sound level of performance with the valid inferences, although there is evidence that the invalid inferences were also supported.

Although only one study of conditional reasoning has been considered here, it is important to note that the findings are typical of the area. The participants here found MP an easier inference to make then MT, and they also showed an inclination towards endorsing fallacious inferences. This is a pattern of performance that can be observed in many studies of conditional reasoning (e.g. Kodroff and Roberge, 1975; Evans, 1977; Wildman and Fletcher, 1977). However, there are no definitive figures which can be placed on the endorsement rates, because reasoning behaviour can be dramatically influenced by the content of the tasks (see Dominowski, 1995), and these vary from one experiment to another. For example some conditionals may invite a biconditional interpretation, and this may lead to an increase in DA and AC inferences being made.

Attention will now be turned to other studies of conditional reasoning in order to provide a good understanding of the conditional reasoning abilities of humans, rather than considering just the results that are found in only one study of the area. Section 1.2.5 will consider how the use of concrete apparatus may aid conditional reasoning performance, if

at all. Chapter 3 will illustrate clearly the influence of content and context on reasoning performance, consequently it is of interest to consider a further study of conditional reasoning which looks at conditional reasoning in a different format to that already considered.

1.2.5 Concrete Apparatus.

The second study that we will now consider is that of Markovits (1988). In this study Markovits examined participants' performance on a conditional reasoning task using concrete apparatus, rather than the traditional paper and pencil tasks that are usually presented to participants. The findings from this novel paradigm will then be compared to results form the traditional reasoning literature note that Legrenzi, 1970 also used concrete apparatus).

It is important here that we should understand the apparatus that Markovits used to examine conditional reasoning in his participants. The piece of equipment used was known as the 'marble game'. This presents the participants with a box on top of which there is a row of open-topped containers. Below the box is another identical row of containers. The participants are informed that hidden tubes link each top container to one bottom container, although it remains unknown as to which top container is linked to which bottom container. The exception to this is a visible tube, which links the two middle containers of each row. The participants are told that "if a marble is put in the top middle container, it will fall into the middle container on the bottom", which is a verbal confirmation of what they can see in front of them. Participants must infer the trajectories

which marbles could take from the top middle container, and the trajectory it could take to the bottom middle container.

Three response patterns were previously noted in participants:

1. Conditional: These are given by participants who consider that the top containers could be linked to any of the bottom ones, including the middle one (i.e. that the relations are not necessarily one-to-one, as clearly indicated by the verbal description given to participants).

2. Intermediate: These responses are mainly given by participants who consider that the top containers could be linked to any of the bottom ones, and (contrary to the verbal description) that the bottom containers could be linked to any of the top ones.

3. Biconditional: participants who consider that the middle container on the bottom is linked only to the middle top container give these responses.

The participants are then shown two concrete manipulations using the apparatus, the first of these is as follows: a marble is placed into the second cup on the top row, and the marble then lands into the second cup on the bottom row. This manipulation was designed to reinforce a one-to-one representation of the apparatus. However, the second of these two manipulations gives more important information to the participants. A marble is dropped into the fifth cup on the top row, and this reappears in the second cup on the bottom row. This was presented to the participants for two reasons, firstly, it clearly shows that two different tubes can lead to the same cup, and secondly that it is possible for tubes to cross inside the apparatus.

Markovits made a number of predictions. Firstly, it was hypothesised that participants would not change their response patterns after they were shown the two experimental manipulations outlined above, even when the evidence from these manipulations was in direct opposition to the previous inferences that participants had made about how the apparatus functioned. Secondly, it was suggested that those participants who altered their reasoning after being shown the first manipulation would tend to give more biconditional responses. They believed that the bottom middle container is linked only to the top middle container. In addition to this, after the second manipulation, it was suggested that there would be an increase in conditional responding: that any of the top containers could be linked to any of the bottom ones. The last hypothesis that Markovits proposed was that intermediate response patterns are a transition between the two relatively stable response patterns represented by conditional and biconditional responses. Thus he argued that participants giving intermediate responses would show the greatest degree of variability in their responses after the manipulations had been shown to the participants.

It is important here to quickly outline the experimental procedure that was used by Markovits in this experiment. The participants were seated near the experimental apparatus, and they were told that if a marble was put into the middle cup (cup 3/ top) on the top, then it will land in the middle cup on the bottom (cup 3/ bottom). They are then shown that this is true by dropping a marble into cup 3 on the top. The subject can see it land into cup 3 on the bottom, this is clear to participants, as the link between these two cups is visible, whereas all the other tubes are invisible to the participants.

The participants were then asked the following four questions:

1. If a marble is put into Cup 3/ top, where could it go?

2. If a marble has landed in Cup 3/ bottom, where could it have come from?

3. If a marble is put into Cup 1/ top, where could it go?

4. If a marble has landed in Cup 1/ bottom, where could it have come from?

After being asked these questions the participants were then presented with the two experimental manipulations that were outlined above, and then asked the four questions again. Response patterns were considered to be conditional if the participants answered all the questions correctly, to be biconditional if the answers they gave indicated that they believed the only way to get a marble into cup 3/ bottom was to go via the visible tube from cup 3/ top. All other response patterns were considered to be intermediate.

The results from these experiments showed that some of the participants did make the changes to their reasoning, as was hypothesised by Markovits. Specifically, participants did show more biconditional responses to the task after the first manipulation, when compared to their initial responses and fewer such responses after the second manipulation. It is interesting to note that 64% of the participants showed stability across the three trials, that is that they did not alter their initial choice to something else.

The results here are also interesting in that they serve to highlight that participants are reticent to change any decisions that they may have made on the basis of incoming

empirical data. Yet, as Evans, Newstead and Byrne (1993) point out, it shows a reasonable level of reasoning competence. Markovits' experiment is interesting in that it allows participants to change their answers after seeing the apparatus work, although the majority chose not to alter their answers.

The results of Markovits' (1988) experiments are in keeping with other conditional reasoning research in that participants display some level of reasoning competence, and yet also make errors, and exhibit evidence of a biconditional interpretation on occasions.

1.3 Theories of Reasoning.

Having considered some of the research into human reasoning, it would seem that humans have some grasp of reasoning, and yet they do make mistakes, and sometimes these mistakes follow a pattern, rather than just occurring randomly. Thus, it is now time to consider some of the theories that have been put forward to explain human reasoning. There are two main theories that have been proposed here, firstly theories of mental logic, these suggest that the mind contains formal inference rules; and secondly, mental models theory as proposed by Johnson-Laird and Byrne (1991). Other theories have also been presented, namely, Heuristic-Analytic theory (Evans, 1984, 1989, 1995, 1996) Optimal Data Selection (Oaksford and Chater, 1994, 1995, 1996) and the Pragmatic Reasoning Schema (PRS) theory of Cheng and Holyoak (1985). PRS theory will be considered in the next chapter, and not here, however, mental logic and optimal data selection will now be considered here.

1.3.1 Mental Logic.

Mental logic theorists, such as Rips (1983, 1994) and Braine and O'Brien (1991), argue that humans use language based representations, and cognitive processes which are similar to natural deduction methods which have been designed by logicians. Thus the basic principle of the approach is that is that the mind contains the various inference rules necessary for humans to reason their way through the world.

As noted by Evans et al. (1993) the various theories differ slightly, however, they are sufficiently similar to be treated here as a theory type, rather than as the individual theories that they are. It should be noted here that the general discussion will consider Braine and O'Brien's (1991) theory, and Rips' (1994) theory separately within the context of the results reported here. Mental logic theories consist of three basic steps to making an inference:

1. Firstly reasoners must convert the premises into an abstract logical form.

2. Secondly, they must retrieve the catalogue of inference rules to construct a mental derivation or proof of a conclusion.

3. Finally, they must convert the content-free conclusion back into the terms of the premises.

For example, the reasoner is presented with the conditional statement: *If Oscar is a cat, then he has a tail,* and given the additional information that *Oscar is a cat.* The reasoner, according to mental logicians, translates this into an abstract form for example:

If p, then q.

p.

Then according to the second step of making an inference, they call-up the content-free rule corresponding to modus ponens, as described above, and construct a proof of the conclusion:

∴ q.

Finally, in following the third step of making an inference, the reasoner must translate this content free logic form back to the form of the original premises, thus they conclude that: *Oscar has a tail.*

Another important aspect to mental logic theories is of relevance here, that is that of the reasoning program. This component of mental logic controls how the rules are selected, modus ponens, in the example given above, and how and when they are applied during the reasoning process, and when to move on to the next step of the proof. Ultimately, as mentioned above, there is the third component that decodes the premises into the logical form, and then re-codes the conclusion back into the form of the premises.

A final point, with regard to the workings of mental logic theories is that of direct and indirect reasoning, as proposed by Brain, Reiser and Rumain (1984), and elaborated by Braine and O'Brien (1991). Direct reasoning occurs when rules are matched to the propositions in the premises: those that can be applied are applied, and each inference is added to the premise set until a conclusion is reached. However, on occasion this direct reasoning may fail, and indirect reasoning strategies are applied, such as heuristics, which will be covered later.

Thus, easy valid inferences are made by the reasoner applying the corresponding inference rule, such as MP. However, this is not the same in the case of difficult valid inferences, as they have no corresponding inference rules, like MT, i.e. there is no one rule in logic for deriving the MT inference. Reasoners must construct a proof through several steps to infer the conclusion. Thus considering the number of steps required and also other factors such as how the reasoning program selects the rules required can assess the difficulty of a problem. Therefore, mental logicians account for the errors that participants make by claiming that they make mistakes on these difficult tasks, by not applying the right rules at the appropriate times (Braine and O'Brien, 1991; O'Brien, 1993). It has also been argued by mental logicians that errors may occur because, although participants may have the rules necessary, they have some difficulty in accessing the rules, and thus the result is that participants make mistakes (Rips and Conrad, 1983). Finally, there is the claim that participants make the errors they do because of a 'failure to accept the logical task' (Henle, 1962). Specifically, participants tend to endorse conclusions that they find believable rather than assessing the validity of them, which is what tasks of logic require participants to do. The outcome of these explanations of errors is that mental logicians claim that it is possible to account for both reasoning competence (the application of rules such as MP), and reasoning errors (failure to accept the logical task, misapplication of rules, failure to access rules).

Finally, some researchers have put forward various other reasons as to why participants show errors on reasoning tasks, such as the comprehension component, which, as mentioned above, decodes the logical form of the premises as they are presented to the participants. (Evans et al, 1993). Participants may also add an additional premise, such as a

common-place assumption that the participants believe to be true in the world, but that is not stated in the premises of the task (Evans, 1993). The addition of premises can cause changes in reasoning behaviour (Byrne, 1989). The above discussion represents the general principles of mental logic theories, it is now important to consider the evidence that exists for mental logic.

Braine et al. (1984), argued that their proposal that the simplest inference problems are those which follow directly from the lexical entry is supported by the low rates of errors on such tasks. In addition to this they argued that those inferences which require the reasoner to make more steps in order to reach a conclusion are more difficult. In order to test this they measured task difficulty by requiring participants to rate the difficulty of the task on a 9-point scale, and combined this with the latencies to solve the problem and the error rates on the problems. They found that task difficulty was a function of the number of steps required and the number of words in the task. The mental logic approach has been criticised however, and these criticisms will be addressed below in section 1.3.2.

1.3.2 Criticisms of Mental Logic.

Firstly mental logic has been criticised by Cheng and Holyoak (1985) for its inability to offer a satisfactory explanation for content effects in reasoning, that is that familiar or concrete materials in tasks can greatly alter the correct response rates on tasks (see Dominowski, 1995 and Chapter 3 here). Mental logicians have argued that content effects can be accounted for in the comprehension component, that is the decoding and re-coding section of the theory, and yet this step of the theory remains remarkably unexplained by mental logic theorists. (Byrne, 1991; Evans, 1989). Both Braine and O'Brien (1991) and

Rips (1994) have suggested that some form of schema theory (such as Cheng and Holyoak's, 1985 Pragmatic Reasoning Schema Theory) could be applied when the content of the task is suitable.

Johnson-Laird, Byrne and Schaeken (1992) made a further criticism of the mental logic view of reasoning. They pointed out that rule theorists cannot explain why Modus Tollens is an easier inference to make from a biconditional, than from a conditional, when the same number of steps is required in both cases. Mental logic theorists have claimed that as the number of steps required to derive a proof increases, so does task difficulty. This idea is disputed by the findings of Johnson-Laird et al. (1992).

Finally, as mentioned above, recent arguments by rule theorists (e.g. Braine and O'Brien, 1991) have suggested that the inference rules of the mind could be supplemented with content-sensitive rules in order to account for some of the data that has been found (e.g. Griggs and Cox, 1982; Johnson-Laird, Legrenzi and Legrenzi, 1972; Wason and Shapiro, 1971). However, this has been criticised by Manktelow and Over (1991) who argued that it weakens the entire theory on grounds of parsimony and testability. It is also unclear as to how this switch from one method of reasoning to another would occur, and what factors in the problem type would initiate such a change in the reasoning process.

Finally, Lowe (1993) has suggested that there is no reason why minds should contain the rules of formal logic as many aspects of formal logic have only recently been developed, and thus to suggest that we have been genetically equipped with such rules is outrageous. However, O'Brien (1995) has responded to this accusation by stating that he does not claim that minds contain formal 'book' logic, but rather abstract rules that concur with logical

proofs. It is important to note here that Lowe is not criticising only formal logic, but also mental models theory (see Johnson-Laird, 1983; Johnson-Laird and Byrne, 1991; and Chapter 2 here). Lowe claims that the concept of humans having logic programs is 'ludicrous' regardless of whether it is in the form of 'natural deduction', or some kind of 'semantic tableau' (Lowe, 1993, p.223).

1.4 Optimal Data Selection

1.4.1 Introduction to the Theory

An interesting recent theoretical development has been that of Optimal Data Selection (ODS) as proposed by Oaksford and Chater (1994, 1995, 1996). The theory is based on Anderson's rational analysis model of cognition (as applied to categorisation, for example) and sees the selection task very differently form other theories of reasoning. Rather than considering the selection task as requiring the participants to engage in logical reasoning, in some form, as other theories do ODS theory considers that the task requires participants to decide between two competing hypotheses. Specifically, these are the dependence model (that p and q occur together: q depends on p) and the independence model (p and q are independent of each other).

Oaksford and Chater have argued that the goal for participants when presented with the selection task is to reduce their uncertainty about the rule (decide between the two models described above). They also argue that the falsificationist view of science (Popper, 1959) is outdated, and no longer the dominant view in the philosophy of science, and thus to expect falsification by participants on the task (via the selection of the p and not-q cards) is unrealistic.

Oaksford and Chater argued that participants reduce their uncertainty about the rule by selecting those cards that will be most informative in relation to reducing uncertainty about the two models. It was argued that initially the two models are seen as equiprobable – they both have a probability of .5. This is then combined with information about the probabilities of p and q, and thus an estimate of the information gain of examining each card can then be estimated. Information gain is the difference in uncertainty about the two models before and after receiving data – turning a card. Thus beliefs held about the probabilities of p and q will affect the estimates about which cards will contain the most information, and this will subsequently influence the cards that are selected.

When presented with the selection task it is impossible for participants to know what is on the unseen side of the cards. Thus information gain must be calculated for all possible alternatives: p and not- p for the q and not cards and q and not-q for the p and not-p cards. All that can be calculated then is *expected* information gain (EIg). Each of the EIg values is then scaled against the average value for each card, in order to reflect the perceived distinctions between cards, yielding scaled expected information gain or SE(Ig). Thus the higher the value of SE(Ig) the more likely the card is to be selected. It is important to note here the importance of what Oaksford and Chater call the rarity assumption. This is the assumption that p and q are rare (in respect to not-p and not-q respectively). When the rarity assumption holds the p and q cards will be the most informative. Subjective probabilities can be calculated using Bayes' theorem.

When SE(Ig) is calculated for the four cards in the standard task the following order card (in terms of SE(Ig))is revealed: p>q>not-q>not-p. This is the standard order of selections

made on the task when presented in it's indicative form (see Chapter Three for details of the selection task literature). Thus it appears that participants are behaving in a rational, though not logical manner. The theory accounts for thematic task performance (where participants do make logical selections) because the rarity assumption is challenged in such cases – p or q may now be seen as common rather then rare. As the occurrences of q increases so does the information gain of selecting the not-q card, and this is the pattern that is observed in versions of the selection task that lead to facilitated performance where more not-q cards are selected than in the standard version (see Chapter three). Thus it is possible for Oaksford and Chater to offer a rational explanation for the observed performance on the selection task in both its indicative and thematic forms. Criticisms of this innovative theoretical approach are considered below:

1.4.2 Criticisms of Optimal Data Selection

The exciting theoretical developments described by Oaksford and Chater (1994, 1996) has attracted a number of criticisms three of which appeared in Psychological review in 1996. These criticisms will now be detailed. Laming (1996) has offered a criticism of the calculations used by Oaksford and Chater, but Oaksford and Chater have successfully rebuffed these critiques. However, Evans and Over (1996a) have offered a more damaging criticism. They argued that equating information gain with reduction (on of the central principles of the theory) is, in itself misguided. They argued that there are occasions where information could be received that would cause a change in information gain, but would not lead to a reduction in uncertainty. For example, if a belief of .25 was held in a hypothesis, then data was received that changed the belief to .75, uncertainty would not have changed (.25 from an absolute value) and yet there would have been information gain.

Thus there is not a direct relationship between information gain and reduction in uncertainty. Oaksford and Chater (1996) have accepted this criticism and altered the way that changes in belief are calculated, but still leaving the theory intact. However, Evans and Over argued that information gain is still too narrow a concept by which to capture changes in belief and argued that epistemic utility is a better measure to use.

Finally Almor and Sloman (1996) argued that the theory is unable to explain the cases where reasoning is facilitated, but that the context is non-deontic, which is sometimes observed. However, Oaksford and Chater can counter this by arguing that any version of the task which overturns the rarity assumption will lead to increased selections of the not-q card. This does not necessarily have to be a deontic version of the task, but any version that challenges the rarity assumption.

Having considered mental logic theories of reasoning and the novel approach of optimal data selection, attention will now be turned to a theory that is more concerned with pragmatic principles than abstract-rule theories. The mental models theory (Johnson-Laird, 1983; Johnson-Laird and Byrne, 1991; Johnson-Laird, 1995 though see also Evans, 1993). This theory will be discussed separately in Chapter 2 because of the importance that this theory will assume later in this thesis.

CHAPTER 2

Page

2.1 A Semantic Theory of Reasoning

2.1.1 Chapter Overview

As has been explained in chapter one in the discussion of inference rule based theories of human reasoning, it is necessary for formal inference rule theories to invoke pragmatic principles in order to account for some of the phenomena that have been observed in empirical studies of human reasoning. However, there are theories in which pragmatic principles are involved more directly: mental models theory (Johnson-Laird, 1983; Johnson-Laird and Byrne, 1991; Johnson-Laird, 1995). Mental models theory is a semantic theory as opposed to the syntactic approach of abstract rule theories.

As with formal inference rule theories in chapter one, the first section of this chapter will be concerned with a discussion of mental models theory. The structure of this chapter will be as follows: firstly it will outline the basic principles of the theory. Then how the theory explains the results of empirical studies of reasoning, along with the evidence that supports the theory. Finally, the criticisms which have been levelled at the theory will be addressed, as well as and how the proponents of the theory have answered these criticisms, including some recent developments and changes to the theory.

2.2 Mental Models Theory

2.2.1. Basic Principles

Attention will now be turned to mental models theory as proposed by Johnson-Laird and Byrne (1991). Johnson-Laird and Byrne characterise reasoning in terms of both what the mind does when it makes a deduction, and how the mind makes a deduction, based on the computational / algorithmic levels of Marr (1982).

In terms of what the mind does Johnson-Laird (1983) describes three constraints that govern the deductions which people draw from premises that they are given. Johnson-Laird states that people's deductions do conform to these constraints, although they may not be conscious of these constraints when making a deduction:

1. People do not throw away semantic information: although a valid deduction does not increase semantic information, participants will avoid those inferences that decrease semantic information.

2. Conclusions should always be parsimonious: people do not draw any conclusions which reassert something which has already been asserted, such trivial conclusions are infinite, and thus consideration of them would lead to an overload of working memory (Baddeley and Hitch, 1974, Baddeley, 1986) very quickly.

3. Any conclusion that is drawn should be informative: it should contain information that is not stated explicitly in the premises.

Having considered what the mind does when making a deduction, attention will now be turned to how the mind makes a deduction. Johnson-Laird and Byrne (1991), propose that the mind does not contain formal inference rules, as proposed by Rips (1983) and Braine and O'Brien, (1991), but rather a set of procedures which are used to manipulate mental models (Johnson-Laird, 1983). By the term mental models Johnson-Laird refers to a mental representation of how the world would be if the premise were true.

2.2.2 The Structure and Manipulation of Mental Models

For example, consider again the conditional *If Oscar is a cat, then he has a tail.* Participants do not represent mentally any cases where the premise is false, and so for a conditional such as this there are three cases where the premise could be true, based on truth tables.

The three cases that people represent are as follows:

1. Oscar is a cat	he has a tail
2. Oscar is not a cat	he has a tail
3. Oscar is not a cat	he does not have a tail

Each of these corresponds to a situation in the world, and each separate model is represented on a separate line. The case in which it is true that Oscar is a cat but false that he has a tail, is false when these two assertions are conjoined by *if.* Therefore, this case is

not represented here, as it is unnecessary (see truth table 1.1 earlier). The theory claims that models will represent as little as possible for reasons of cognitive economy.

However, it is of some discussion as to exactly what the nature of these models is, that is, how are they represented in the mind? They may be images, or they may be propositional tags (see Pylyshyn, 1973 and Kosslyn, 1994, for a detailed discussion of the imagery debate). However, the nature of the models is not the critical point within the theory. What is crucial, however, is the structure of the models, and how humans manipulate them in order to make deductions.

In order to reason about the conditional here 'c' will be used to represent: 'Oscar is a cat', and 't' will be used to represent: 'he has a tail', and '¬' is used to represent negation, such that '¬t' means 'does not have a tail'. This is the notation used by Johnson-Laird and Byrne (1991). Therefore, using this notation, the models that would be constructed, according to mental models theory would be as follows:

c	t	(Is a cat, has a tail)
¬c	t	(Is not a cat, has a tail)
¬c	¬t	(Is not a cat, does not have a tail)

Further to this, as was mentioned earlier, working memory limitations play a role in the way that models are constructed (e.g. Quayle and Ball, 1997). As a result of this, as much information as possible is represented *implicitly* in the model, for example:

c	t	(Is a cat, has a tail)
...		

Clearly, the top line represents 'is a cat' and 'has a tail', the three dots in the bottom line are used to express that there are further models possible, the content of which is not yet explicit (Johnson-Laird and Byrne, 1991). In addition to this there is also the use of square brackets, for example:

[c] t

This is used to indicate exhaustive representation of all instances of this contingency: there cannot be a case in which Oscar is a cat, without his having a tail (this would be the case from truth tables which result in a false outcome when the two assertions are conjoined by *if*, thus there is only one model which contains c as affirmative). However, the information that Oscar has a tail has not been exhaustively represented (thus it has not been expressed in square brackets), and so this may occur in other models. When these other models (represented by the three dots above, in the initial model), are fleshed out, they are as follows:

c t

¬c t

¬c ¬t

The above section has considered how mental models are constructed, and the information they represent, however, it is the manipulation of models which allows reasoning to occur. The theory of mental models now explains how these models are manipulated in order to make the deduction necessary. The model theory of propositional reasoning specifies that reasoning requires 3 stages (Johnson-Laird and Byrne, 1991).

1. The first stage of this reasoning involves model constructing procedures. These procedures take the information in the premises and convert them into the model format, they also add to the premises any knowledge of the area that the person may already have. Finally, the model constructing procedure 'fleshes out' the models to be more explicit, if this proves to be necessary.

2. The second stage of the reasoning process involves the model combining procedures, these procedures take two or more sets of models and combine them together, revising them as they go in order to remove any inconsistencies. This second step of the process also has model describing procedures, which produce a parsimonious description of the models.

3.The final stage of the reasoning process, according to Johnson-Laird and Byrne, is the validation process. This is where the conclusion, if one is possible, is produced. The validation process involves taking a set of models and a putative conclusion, and tries to produce a new set of models, and a conclusion that falsifies the putative conclusion. New models are produced which are consistent with the premises, but which may be inconsistent with the putative conclusion that has been drawn, and if this is the case then the putative conclusion is rejected, and a new conclusion is sought.

The theory of mental models suggests that humans have the abilities to make rational deductions, and yet they do make errors when applying the general principles of deduction (Johnson-Laird and Byrne, 1993). Johnson-Laird and Byrne term this as "rational in principle, but erring in practice." This is how the mental models theory can account for the

pattern of responding that was presented in Chapter One - people have the tools to make deductions, but they make mistakes when using these tools.

For example modus ponens is a relatively easy inference to make, and humans show little difficulty in making this inference (as discussed earlier e.g. Wildman and Fletcher, 1977). However, modus tollens is a harder inference for humans to make, and this is borne out in the performance of humans when presented with tasks that require them to make this inference. They are more likely to show correct performance when making a modus ponens inference than when they are making a modus tollens inference. For example consider the models required to draw the modus ponens inference:

Given the following:

If Oscar is a cat then he has a tail

Oscar is a cat

Following the principles of the theory as outlined above the first stage of the theory is the model constructing process, thus an initial model set is constructed which corresponds to the first premise above:

[c] t

...

Where 'c' corresponds to 'is a cat', and 't' to 'has a tail' and '...' corresponds to other possible models, which as yet have no explicit content.

There is also a model corresponding to the second premise:

c

Thus the first stage of the process is complete, and the second stage begins, this is the model combining procedures. During this process the information from the second model is added to the first model in the initial set, and the implicit model is eliminated, due to exhaustive nature of the first part of the first model. As mentioned above this stage also features model describing processes, and the model which results could be described as:

Oscar is a cat and has a tail.

However, the processes of making a modus tollens inference are not quite so straightforward, given the following:

If Oscar is a cat then he has a tail

Oscar does not have a tail

As with Modus ponens, reasoners represent the first premise as:

[c] t

...

The second premise is represented as:

¬t

However, unlike modus ponens the information in these two premises cannot be combined here, and so some reasoners conclude that nothing follows from the premises. If the set of implicit models (represented in the initial model set by '...') are fully 'fleshed out' they are as follows:

c t

¬c t

¬c ¬t

The second stage of reasoning, combining the models, now adds the information from the second premise to the information in the third model set above. This now eliminates the first two models, as they contain information about Oscar having a tail, and this leaves the final model:

¬c ¬t

The procedures that describe the models now produce the parsimonious conclusion that Oscar is not a cat.

Thus it is clear that modus tollens is a far harder conclusion to draw than modus ponens, because models must be fleshed out fully (made explicit) and multiple models need to be kept in mind at the same time. From the above examples it can be seen that mental models theory is distinct from theories of formal inference rules, as rules are not applied in model theory, rather the models are constructed and manipulated in order to produce conclusions.

2.3 Evidence for the Mental Models Theory

2.3.1 Model Theory and Reasoning Errors.

Model theory can also account for why humans make some of the mistakes they do when they are presented with tasks that require them to draw inferences. Errors may occur if

reasoners do not flesh out the initially implicit models fully. This was illustrated above with modus tollens. Errors may also occur due to failure to consider a possible model to be so, or due to working memory limitations: the inability to be able to keep all the relevant models in mind at any one time. (Baddeley, 1986). Two classic errors that participants make are denial of the antecedent (DA), and affirmation of the consequent (AC). These errors are as follows. Consider again the conditional, *If Oscar is a cat then he has a tail*, denial of the antecedent is committed if when given the second premise *Oscar is not a cat*, they draw the conclusion that *Oscar does not have a tail.* The affirmation of the consequent error is made when humans are given the second premise *Oscar has a tail,* and they draw the fallacious conclusion that *Oscar is a cat,* as discussed earlier.

These errors can occur if the participants consider that there are no alternatives to Oscar being a cat, which could result in his having a tail (being a dog, mouse, horse etc.). Logically, participants take the statement as being a biconditional rather than a conditional (or expressing material equivalence rather than material implication). In terms of our statement above, the biconditional would be *if and only if Oscar is a cat then he has a tail.* As with both the MP and MT inferences the humans construct an initial model set, however, in this case both components of the model are exhaustively represented:

[c] [t]

...

Given then that Oscar has a tail, the procedures used for combining the models, can eliminate the implicit models above, and the conclusion is drawn that Oscar is a cat, and thus the AC error is committed, although this is not an error given a material equivalence (biconditional) reading.

In terms of the DA error participants are given the second premise that Oscar is not a cat, they may initially draw no conclusion because they are unable to combine the information in the second premise with that of the initial model set. As above, their initial model set will be:

[c] [t]

...

Thus the model for the second premise, that Oscar is not a cat:

¬c

is not able to be combined with the models above, however, if the implicit models above are fleshed out then the full model set for the biconditional interpretation is:

c t

¬c ¬t

It is then possible for the information from the second premise to be combined with this model set, and the first model is then eliminated, and the DA conclusion that Oscar does not have a tail is subsequently drawn.

Thus mental models theory is able to account for the four inferences that have been observed in studies of conditional reasoning, that is MP (Modus Ponens), MT (Modus Tollens), AC (Affirmation of the Consequent) and DC (Denial of the Consequent).

In addition to being able to account for the observed patterns of performance, there is further evidence for mental models theory that has been offered by Johnson-Laird and Byrne (1991). As has been mentioned before, the greater the number of explicit models that humans need to consider for any given deduction, the harder the inference is. In

addition to this any deduction that can be made by using only the initial models is easier than any deduction which requires the models to be fleshed out to an explicit level. Finally, Johnson-Laird and Byrne (1993) point out that it takes time to check for inconsistencies between models.

As we have seen above, mental models theory can account for the differences in performance between modus ponens and modus tollens for a conditional. It can also account for the lack of difference between MT and MP for a biconditional, in that the initial set of models contains explicit information about both the affirmative and negative cases. Using the example above, in its biconditional form 'Oscar is a cat only if he has a tail' yields the following set of initial models:

[c] t

¬c [¬t]

This initial set of models allow both the MP and MT inferences to be made without the need for any further fleshing out. However, because a biconditional requires two models MP should be harder to draw under a biconditional interpretation. Johnson-Laird and Byrne (1991) claimed that the data support this assertion.

The theory can account for the ambiguity sometimes shown by reasoners as to whether 'if' is to be interpreted as a conditional or a biconditional, because as we have seen above initial models can be fleshed out to be either. Finally mental models theory can also account for the drawing or endorsing of fallacious inferences under a conditional interpretation (e.g. Evans, 1977; Wildman and Fletcher, 1977).

2.4 Criticisms of Mental Models.

Attention will now be turned to some of the criticisms that have been made of mental models theory. Initially empirical criticisms will be made, and then criticisms of a theoretical nature will be examined.

Some researchers have failed to find support for the predictions of the theory. Rips (1990), claimed to have found no difference in the correct deductions made by participants regardless of whether they were single model problems, or multiple model problems (approximately 65% correct for both type of problem). As was seen above with the MT inference example, multiple model problems should be harder to solve. However, Rips' theory predicts that there should be no difference in performance rates on problems involving the negation of a conjunction, and problems involving the negation of a disjunction. Whereas mental models theory would predict that there would be a difference, due to the negation of a conjunction leading to three models, and the negation of a disjunction leading to only a single model. Byrne and Handley (1992), however, showed that reasoners rarely made the correct interpretation of the three model problem (25%), whereas performance on the single model problem was better (43%), thus offering support for the predictions of mental models theory.

Attention will now be turned to some theoretical criticisms. Rips (1986) argued that although the theory could explain content effects (something that poses difficulty for formal rule theorists, as was discussed in Chapter 1), it was unable to explain reasoning with unfamiliar material. However, Johnson-Laird and Byrne, (1991) replied to this criticism by

suggesting that humans could reason successfully with unfamiliar materials because all they need is knowledge of the connectives (if, and, or etc.) and the quantifiers (some, many, none etc.), rather than specific knowledge regarding the materials.

Green (1993) and Eysenck and Keane (1990) have criticised the theory on the grounds that the process by which world knowledge is incorporated into the models that people use, and how people retrieve counterexamples, remains unclear.

It is the search for counterexamples, which is a central issue of mental models theory, in that a putative conclusion will be tested by the search for counterexamples. However, this element of the theory appears to motivate few predictions that are made based on the theory (Polk, 1993). Bara (1993) and Fetzer (1993) have argued that searching for counterexamples is a very difficult task for humans to attempt successfully. Finally, Rips (1990) contends that the theory for propositional reasoning appears to rest on an algorithm that does not involve the search for counterexamples.

However, in reply to this Johnson-Laird and Byrne (1993) suggest that the search for counterexamples is not necessarily a self-conscious one, and reasoners frequently make mistakes at this stage, or may fail to complete it at all. In response to the search for counterexamples being difficult, Johnson-Laird and Byrne argue that it is this very difficulty which results in the prediction that the greater the number of models in the deduction, the more difficult the inference will be. However it should be noted that this does appear to be a slightly different argument from saying that it is the number of models which need to be considered that make an inference difficult.

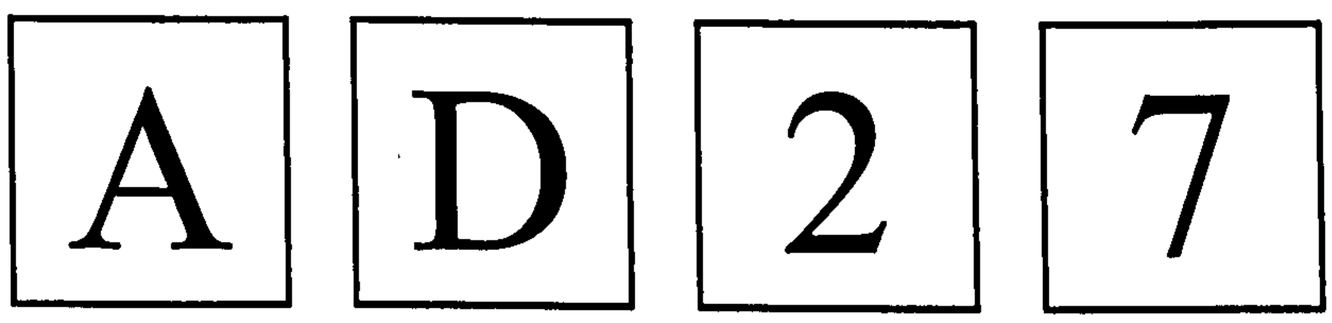
One of the most commonly used tasks for studying conditional reasoning in humans over the past thirty years has been the Wason selection task, and it is this task that is the sole tool used here for research purposes. The task, along with some of the results that its use has yielded, and the various theories used to explain these results will be considered in detail in the next chapter. However, the task will be described briefly here, along with the mental models account of task performance, and its claim to account for the 'five phenomena' of the selection task.

2.5 The Selection Task and Mental Models Theory

2.5.1 The Selection Task

The basic selection task (Wason, 1966, 1968) is as follows. A participant is presented with four cards such as those given in Figure 2.1 below, and is also given the rule '*If there is an A on one side, then there is a 2 on the other side*'

Figure 2.1



Cards used in the selection task (Wason, 1966)

Participants are required to select those cards and only those cards that would test whether the rule was true or false. The correct solution to the task is to select the 'A' and '7' cards, as these are the only cards that could contain the falsifying combination of an 'A' with a number other than '2'. The task is discussed in more detail in section 3.2.1 below.

The aim of this section is to briefly discuss the selection task in terms of the mental models theory. The theory is also considered in detail in the General Discussion, and the analysis here will be brief as a result.

Johnson-Laird and Byrne (1991) list the 5 phenomena of the selection task as those manipulations of the task that lead to improved performance on the task. The theory also claims to be able to account for these five factors. These five phenomena (which are discussed in detail in the next chapter) are:

1. Change the form of the rule: matching bias is an example of this facilitatory manipulation.
2. Change the content of the rule: this has been the source of much research on the selection task (e.g. Johnson-Laird, Legrenzi and Legrenzi, 1972; Dominowski, 1995).
3. Change the context of the rule: such as the use of a deontic framework (deontic reasoning is discussed in more detail later). For example, Cheng and Holyoak, (1985), Cosmides, (1989).
4. Change the content of the cards: Jackson and Griggs (1990) found that labelling negative *explicitly* led to improved performance.
5. Altering the task so that participants are more likely to represent all alternatives explicitly in their models. This manipulation includes the RAST (Reduced Array Selection Task) (Wason and Green, 1984), instructing participants to check for violations of the rule (Chrostowski and Griggs, 1985), and making participants verbalise the reasoning behind their selections (Berry, 1983).

Mental models theory claims that it accounts for all of these facilitatory manipulations, and theories that rely on memory for counterexamples, or pragmatic reasoning schemas are

unable to do so. The selection task and examples of these 'five phenomena' will now be considered in detail in the next chapter.

CHAPTER 3

Page

3.1 The Wason Selection Task.

3.1.1 Chapter Overview

The primary tool for the study of conditional reasoning has been the Wason selection task (Wason, 1966, 1968), and the numerous permutations of it which followed over the next thirty years (e.g. Wason and Shapiro, 1971; Evans and Lynch, 1973; Griggs and Cox, 1982; Wason and Green, 1984; Manktelow and Over, 1991; Kirby, 1994). The task has been considered briefly in Chapter two in terms of the mental models account, but the task will be dealt with in detail in this chapter. The task will be considered in detail as it is the tool used in the present research. Following this history of the selection task there will be a short section presenting two domain specific theories which have been proposed to account for the selection task in only some of its forms. This is in contrast to the mental logic and mental models theories considered thus far, which attempt to account for all reasoning with conditionals.

3.2 The Selection Task

3.2.1 The Standard Selection Task

Initial attention will be turned to describing the original task itself, and then some of the numerous different versions that have been presented to participants since the original task. (See Evans, Newstead and Byrne, 1993; and Dominowski, 1995, for a review of the selection task literature). The standard selection task, (Wason, 1966, 1968) or abstract task as it has become known was presented to participants as follows:

A participant is told that a set of cards has been produced all of which have a letter on one side and a number on the other. The participant was then presented with four cards, 2 of which were letter side up, and 2 of which were number side up (see Fig. 3.1). The participant was told that the following rule applies to these four cards, and that it may be true or false: '*If there is an A on one side, then there is a 2 on the other side*'. (In purely logical terms the conditional would be stated as *If p then q*, and the cards would show the four possible values of p, not-p, q and not-q). The task for the participants is to indicate those cards and *only* those cards that need to be turned over in order to decide whether the rule is true or false.

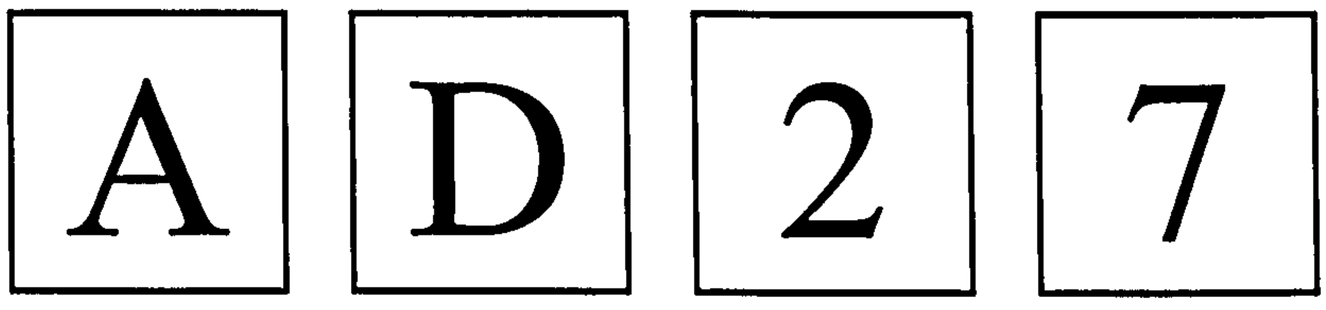


Figure 3.1 Cards that may be used in the selection task.

Although the task was thought to be quite simple, it was clear that many participants had a great deal of difficulty in selecting the correct cards. The most common choices for participants to make were to select the p (A) and q (2) cards, or to select the p (A) card alone (these two selection patterns accounted for 79% of choices made by participants, when presented with the selection task by Johnson-Laird and Wason, (1970), when pooling across four experiments, 46% selecting p,q and 33% selecting the p card alone). These rates are robust and replicated many times e.g. see Dominowski (1995). However, the correct choice for participants to make was to select the p (A) and not-q (7) cards, as only these cards could contain the falsifying combination of an A with a number other than 2. Correct performance on this task is surprisingly low, in

relatively educated participants (usually undergraduates), with only 4% of participants selecting this logically correct combination in Johnson-Laird and Wason's 1970 experiments (again pooled across four experiments).

When participants are presented with the selection task what are they required to do? Evans, (1982) suggested that there are three steps necessary to solve the task:

(i) The participant must appreciate the falsification principle (Popper, 1959). This alludes to the importance of the not-q card, and its potential to falsify the rule, and thus to solve the task correctly.

(ii) The participant must use a truth table in order to determine which cases would be falsifying.

(iii) The participant must decide which of the four cards could reveal a case such as this.

Given that the majority of participants fail to solve the task correctly, they must be failing to complete one or all of these steps. Originally Wason (1966) suggested that participants were trying to verify the rule rather than falsify it, as they should do, hence the p and q card selections. Verification does not allow the participant to state whether the rule is true or false, as even if all four cards with which the participant is presented conform to the rule it still may not be true. However, if only *one* card contains the falsifying combination of an A and a number other than 2, then the participant can state that the rule is false. Thus a series of experiments was devised where participants were given hints or tips towards the importance of the not-q card in the task.

3.2.2 The Therapy Experiments

Wason (1968) presented participants with the task, and they were given clues as to the importance of the falsifying potential of the not-q card (the card which is typically ignored). These experiments, perhaps not surprisingly, have become known as the therapy experiments. Participants have little difficulty in determining possible falsifying situations, this is an expected finding given that participants can display a reasonable knowledge of logical 'truth tables'. As was discussed in Chapter One, truth tables are used by logicians to explain the way a logical operator acts on any given proposition (Johnson-Laird and Tagart, 1969). In addition to this participants were aware of the importance of the not-q card as a card that could reveal such a situation because that is what the therapy experiments were for, to alert the participant to the not-q card. Evans (1982) then suggested that the difficulty for participants may lie in step (i) above, that participants were failing to appreciate the falsification principle. Although note that manipulations in which the participants were explicitly informed that they should attempt to prove the rule false do not lead to an improved rate of performance on the task (Wason and Golding, 1974). However, many participants who had been through this logical therapy session still selected the p and q cards, rather than the potentially falsifying not-q card. The reason for this tendency may lie in a phenomenon termed matching bias (Evans and Lynch, 1973), this factor was revealed through the use of the negations paradigm.

3.2.3 The Negation Paradigm and Matching Bias

Evans and Lynch (1973) presented participants with selection tasks with rules that differed from that used in the standard selection task. These rules saw the introduction of negation. For example, rather than the standard *If p then q*, the rule used would be *If p then not-q*. All four permutations of the rule were used, the two previously mentioned, plus *If not-p then q* and *If not-p then not-q*. The logic here remains the same, although the *values* of the cards which the participants should select alters. In order to explain this fully the cards must now be given a new title, in order to clarify the logical status of the cards: true antecedent (TA), false antecedent (FA), true consequent (TC), false consequent (FC). For the standard task these are clear, however, for the rule *If p then not-q* the TC card is now represented by the not-q card and FC by the q card etc.

Whatever the form of the rule, participants should always select the true antecedent and the false consequent cards, as this is the potentially falsifying combination. Manipulation of the task with negatives allowed Evans and Lynch to test Wason's original hypothesis that participants select the TA and TC cards because they are potential verifiers of the rule (Wason had previously found support for a verification bias in inductive reasoning with the 2-4-6 task, Wason, 1960). The findings from Evans and Lynch suggested that participants were not trying to verify the rule they had been given. Rather they appeared to select those cards which had been mentioned in the rule itself, or to 'match' their choices with the values from the rule hence the term matching bias. Evans and Lynch's findings are summarised below:

Table 3.1 Matching Bias Data from Evans and Lynch, 1973

RULE	TA	FA	TC	FC
(1) If p then q	**88%**	8%	**50%**	33%
(2) If p then not-q	**92%**	4%	8%	**58%**
(3) If not-p then q	58%	**29%**	**58%**	42%
(4) If not-p then not-q	54%	**46%**	29%	**75%**

Evans and Lynch (1973).

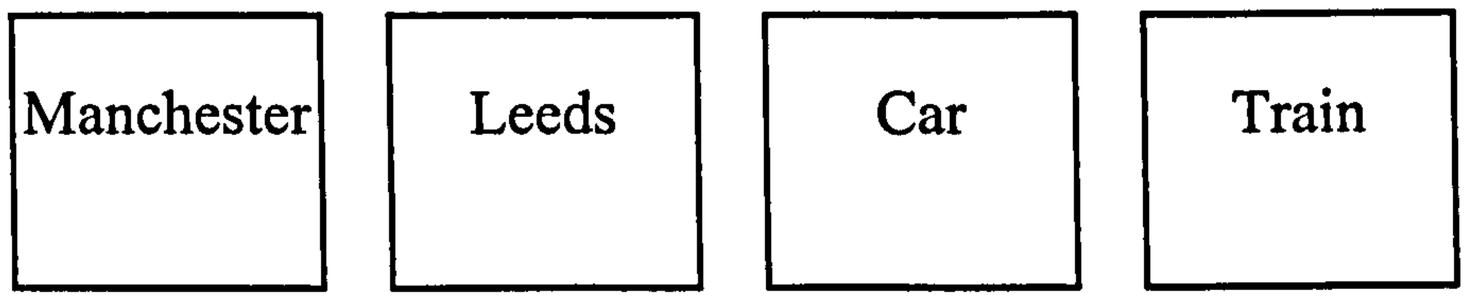
It can be seen from table 3.1 that there is a tendency for participants to select those cases which match the items mentioned in the rule (matching cases are given in **bold**). However, it should be noted that double matching cases are rare, occurring only in the second rule in the table. It appears then that matching occurs more in the consequent cards than in the antecedent cards. The matching bias effect is robust and has been replicated many times, e.g., Manktelow and Evans, (1979), Oaksford and Stenning, (1992). (See Evans, 1998 for a full review of matching bias research).

3.2.4 The Effect of Realistic Content

The selection task data that have been considered so far have concentrated solely on the abstract or standard selection task, using numbers and letters as the card values. Later research revealed that the use of realistic or thematic contents could facilitate participants' performance on the selection task. The first reported case of this comes from a manipulation of the task presented by Wason and Shapiro, (1971, experiment 2). In this version of the task participants were given a claim made by the experimenter: *'Every time I go to Manchester I travel by car.'* Participants were then presented with four

possible values on cards (as in the standard task, examples of the cards are given below in figure 3.2), and asked which of the cards would need to be turned over in order to assess whether the experimenter's claim was true or false.

Figure 3.2 Cards used by Wason and Shapiro (1971)



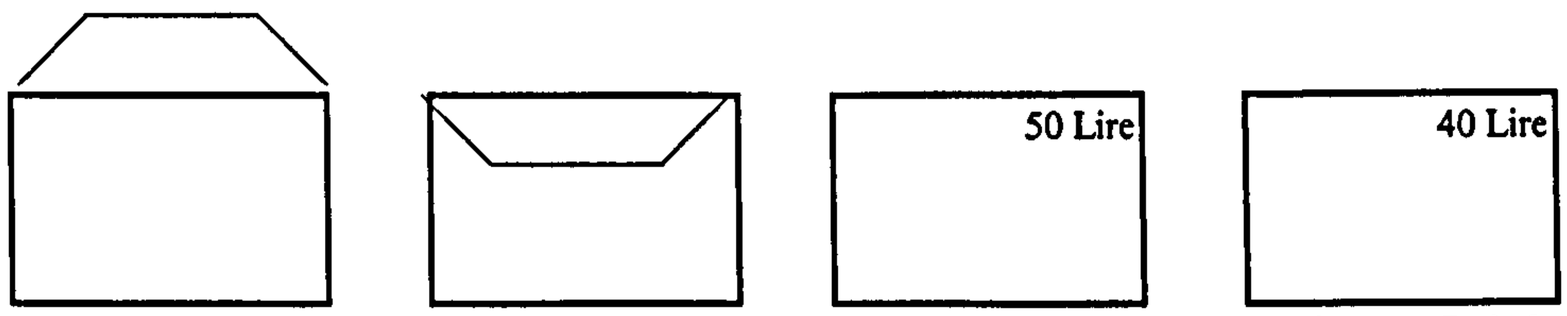
Wason and Shapiro (1971).

When participants were presented with this version of the task performance was dramatically increased when compared to the standard task described in section 3.2.1. In this revised version of the task 63% of participants correctly selected the p and not-q cards (Manchester and Train), on the thematic version of the task, and only 12.5% selected the p and not-q cards on the standard version, in line with previous abstract task performance (Wason, 1966, 1968). It should be noted here that in logical terms the two tasks were seen as isomorphic, with the same logical process required to take place in order for participants to correctly solve the task. Hence it appears that it was the content of the problem that influenced the participants' selections on the task.

Further support for this 'thematic facilitation effect' was found by Johnson-Laird, Legrenzi and Legrenzi, (1972) in a task now known as the postal task. In this task participants were given a rule regarding postal regulations, required to imagine that they were post office workers, and shown four envelopes rather than four cards. The rule they were told was 'If a letter is sealed, then it has a 50 lire stamp on it', and the four

envelopes were as follows: sealed (p), unsealed (not-p), 50 lire (q), and 40 lire (not-q), see figure 3.3 below. Participants were told to turn over those envelopes, and only those envelopes, that they thought they needed to check in order to make sure that the rule was not being broken.

Figure 3.3 Cards Used by Johnson-Laird et al. (1972).



Johnson-Laird, et al. (1972).

As with the Towns and Transport version of the task, presented by Wason and Shapiro, this thematic version of the task produced improved performance, compared to the standard selection task. Over 91% of participants correctly solved this postal version of the selection task, by selecting the 40 lire stamped envelope (not-q) and the sealed envelope (p) for inspection. It is important to note here that the participants used in this experiment were British students who would be familiar with a rule such as this (a postal rule of this nature had previously been in force in Britain). Consequently, the experiment can be criticised on the grounds that it taps directly into knowledge that the participants would have, and thus they bring extra information to their reasoning rather than just their reasoning ability *per se*. It would be possible for the participant to solve the task without engaging in any reasoning at all. Indeed Cheng and Holyoak (1985) repeated this version of the task with participants who were not familiar with such a rule, and they showed poor performance on the task, suggesting that it does rely on some form of recall.

It appeared to be the case that content had a dramatic effect on the performance of participants presented with the selection task. When the task was in its standard abstract form, performance was very poor. When in a thematic, realistic form performance was dramatically improved with participants selecting the logically correct p and not-q cards. Obviously it was of interest that human reasoning performance seemed to be so strictly tied to the content of the problems with which participants were presented. It did question Piaget & Inhelder's (1969) theory of formal operations, where adolescents should acquire the rules of formal logic to guide their thought processes. However, were this the case content should not influence reasoning, as logic should be content independent, and not influenced by the *meaning* of the words in the rule.

However, later studies revealed that the thematic facilitation effect may not be as robust as originally suggested. A series of experiments by Manktelow and Evans (1979) failed to find any facilitatory effect with thematic materials. Manktelow and Evans (1979) presented participants with a range of conditionals concerned with food and drink such as, 'If I eat haddock then I drink gin' and 'If I eat macaroni, then I do not drink champagne.' Using these rules Manktelow and Evans found no difference between these thematic versions of the task and abstract versions using letters and numbers. In addition to this Manktelow and Evans also found no facilitation on the task using Wason and Shapiro's (1971) Towns and Transport version, as presented in figure 3.2. This evidence resulted in Manktelow and Evans claiming that the facilitation effect was not as strong as had originally been suggested from earlier results with thematic materials. However, it should be noted that although the versions used by Manktelow

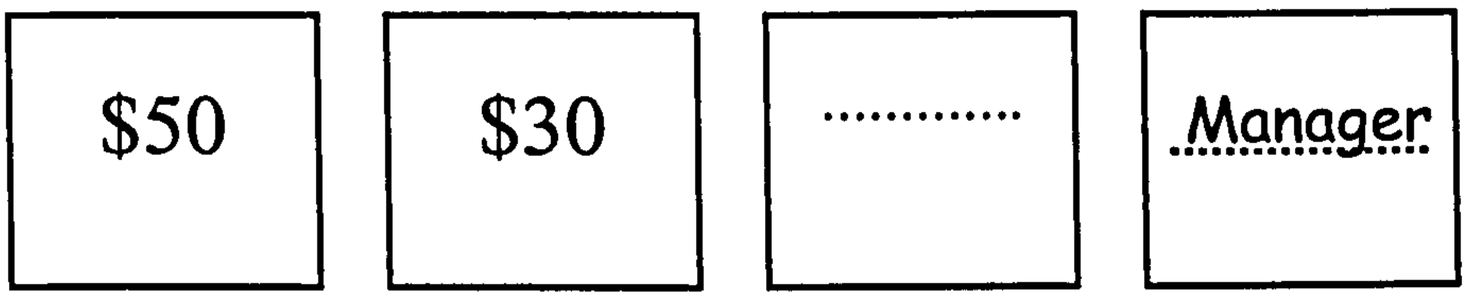
and Evans are thematic in that they are about something real, they are not rules that are common in their usage. Perhaps it is the reverse of the example used by Johnson-Laird et al. (1972) which was criticised for being *too* realistic, with Manktelow and Evans' version not being realistic *enough*.

In summary of the research on the selection task thus far: Initially it appeared that realistic content on the selection task would lead to improved performance, (as had previously been noted by Wilkins, 1928, with syllogisms) and the avoidance of matching bias (Wason and Shapiro, 1971, and Johnson-Laird et al, 1972). However the series of experiments by Manktelow and Evans, (1979) cast doubt on these findings and suggested that the effect of thematic content may not be as robust as had been originally suggested. There was clearly a need for further research to examine the role of content in reasoning, and to tease out those factors that lead to facilitation on the task. Such research was also necessary to enable the findings about thematic material to be accounted for within a theoretical framework. One of the theoretical suggestions of the time was memory cueing, the idea that the participant must be familiar with the rule in order to correctly solve the task, as seemed to be the case with Johnson-Laird et al's (1972) findings, and Manktelow and Evans' results in 1979. However, this theory was soon to be challenged by further work within the influence of content on selection task performance.

In 1980 Rumelhart reported D'Andrade's Sears task. This version of the task has participants showing facilitation although they do not necessarily have any familiarity with the rule as such, thus casting doubt on the idea of memory cueing. This rule was presented to participants as follows: Participants were required to imagine that they

worked in a department store, and were told that 'If a purchase exceeds $30 then the receipt must be approved by the departmental manager'. The four cards that were presented with this task are given below in figure 3.4:

Figure 3.4 Cards used in D'Andrade's Sears Task (as Reported by Rumelhart, 1980).



Rumelhart (1980).

Few people have any difficulty in selecting the p and ¬q cards on this task, that is the receipt for more than $30 and the unsigned receipt. Therefore, as mentioned above the initial idea of memory cueing as the explanation for facilitation was not enough to explain the results, that is participants did not recall the correct solution from previous experience. They had not all worked in department stores where this rule had been enforced, and thus could not be accessing the information from some long-term memory store. Thus they were not solving the task without reasoning about it, as had been claimed of Johnson-Laird et al.'s (1972) findings with the postal task. There must be some other process taking place, some *reasoning* process. Before moving to theoretical explanations of selection task performance, there is one last finding of interest: that of transfer effects.

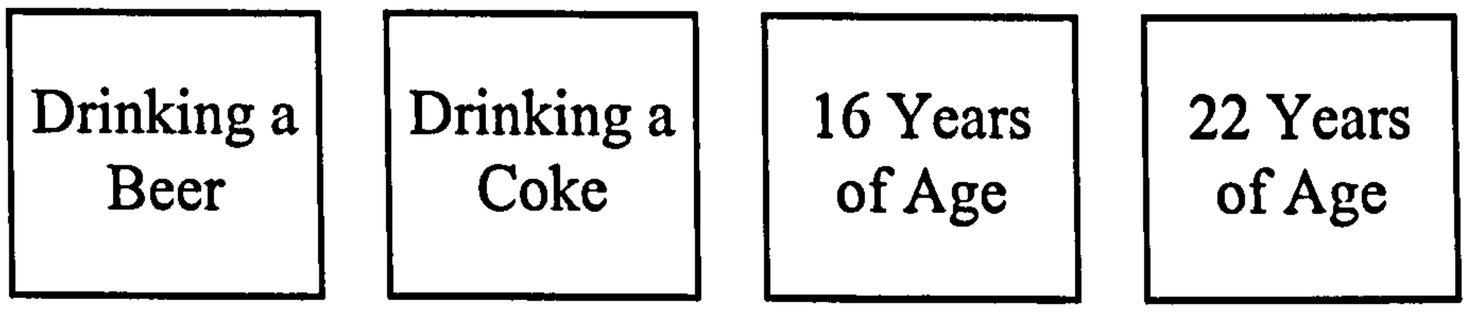
3.2.5. Transfer Effects

Some researchers (Griggs and Cox, 1982) have shown that the facilitation effect can transfer from a task where facilitation is found onto performance on a non-facilitatory

version of the selection task, although these data are somewhat inconclusive. Unlike earlier efforts (Johnson-Laird et al., 1972 and Wason and Shapiro, 1971), Griggs and Cox (1982) were successful in their attempt to show a transfer of facilitation. They used two main rules here, these were the Drinking age rule and the Clothing age rule. The Drinking age rule is as follows:

> On this task imagine that you are a police officer on duty. It is your job to ensure that people conform with certain rules. The cards in front of you have information about four people sitting at a table. On one side of a card is a person's age and on the other side of the card is what a person is drinking. Here is a rule: "IF A PERSON IS DRINKING BEER, THEN THE PERSON MUST BE OVER 19 YEARS OF AGE". Select the card, or cards that you definitely need to turn over to determine whether or not people are violating the rule.

Figure 3.5 Cards Used by Griggs and Cox (1982).



Griggs and Cox (1982).

Participants showed very little difficulty in making the correct selections when presented with this version of the selection task. Facilitation on this task was very strong with 74% of participants checking the beer drinker and the person under 19 years of age, that is the p and ¬q cards (Griggs and Cox, 1982), when compared to standard selection task performance.

However, to return to the main point here, that of transfer of the facilitation effects. Griggs and Cox (1982), did find evidence of facilitation transfer, by using three versions of the task, the drinking age rule as shown above, an abstract version (letters and numbers, as described earlier in section 3.2.1) and the Clothing Age rule. This version of the task has the same set-up as the Drinking Age rule, but the participants are given the following rule: "If a person is wearing blue, then the person must be over 19 years of age". Although this rule has the same structure as the Drinking Age rule, it has no real world reference, that is, that there is no age limit as to what colour of clothes people can wear in reality, whereas there is an age limit to drinking alcohol in a public place. The important factor in participants making the correct selections on the Clothing Age rule is the order in which the three tasks where presented to the participants. When the Clothing Age rule was presented first there was no facilitation effect. However, when the Drinking Age rule was presented first and the Clothing Age rule followed there was a significant transfer effect. However, there was no transfer to the abstract letters and numbers version of the task. Thus it appears that memory cueing is not responsible for facilitation effects. It is the *deontic* content of the rules. This will now be discussed in section 3.3.1.

3.3 Deontic Reasoning

3.3.1 Deontic Content

The important factor here which was initially pointed out by Cheng and Holyoak (1985) is that in the Drinking Age Rule task the participants are required to check for rule violation, whereas in the Johnson-Laird, Legrenzi and Legrenzi task the participants were claim testing, rather than checking for violations. Thus in the Postal task the participants rely on memory cueing for facilitation (the participants who were

familiar with such a rule show the facilitation), whereas the Drinking Age rule does not, and participants who have no experience of such a rule can solve the task. The important criterion then for facilitation is that the rule should be in a deontic form. Deontic reasoning is a form of reasoning that is concerned with actions that should, may or ought to be performed. Over and Manktelow (1993) defined deontic reasoning as part of what philosophers term practical reasoning, and is distinct from theoretical reasoning. Theoretical reasoning is concerned with discovering objective matters of fact. Alternatively practical reasoning is concerned with inferring what we should, may or ought to *do,* it is concerned with actions. However, it is true that theoretical reasoning may have implications for practical reasoning, though the latter goes further to make decisions about actions. Given that deontic reasoning is reasoning about actions it is somewhat strange that psychologists ignored it for so long, as it is a form of reasoning in which we must regularly engage. As mentioned above it was Cheng and Holyoak who first claimed that it was a deontic context that would yield facilitation on the selection task. Specifically, facilitation on a problem may be improved if the task involves the search for violations of the rule (search for violations is not sufficient for facilitation - though it may be necessary), rather than just testing the rule, as was the case in the standard task, and the rule should contain modal terms like 'may' or 'must'.

As well as pointing out the importance of the deontic context Cheng and Holyoak also proposed a theoretical account of performance on the selection task, which was domain specific, in that it only relates to the selection task when it is presented within a particular form, its deontic form. Below will be a brief summary of their research, and then Cheng and Holyoak's (1985) theory will be presented.

3.4 Domain Specific Theories

3.4.1 Pragmatic Reasoning Schema Theory

As was mentioned above, the selection task has been the primary tool used in reasoning research over the past thirty years, and it has appeared in many different forms, some of which have been presented here. There are also theories which have been developed primarily to explain performance on the selection task, particularly thematic versions of the task, such as the Drinking Age rule (Griggs and Cox, 1982). The main theory here, which has been the motivation for much of the recent research in the area of deontic reasoning, is that of Pragmatic Reasoning Schemas (PRS), as proposed by Cheng and Holyoak (1985). PRS theory is a domain specific theory that attempts to account for facilitation on the selection task. Before the theory is considered in detail the concept of schemas generally must be explained.

Thus, it is necessary that schemas are fully understood, as these underpin the theory itself. Schemas are similar to what would be generally called concepts, but yet, they can be more specific than concepts, in that they can actually contain rules for inference. However, schemas contain knowledge that has been abstracted, or drawn from prior experience in a particular field or area (see Anderson, 1995 for a discussion of the role of schemas in cognition). Thus, schemas can be used to explain the robust facilitation effect that has been observed when the task has been presented to participants with a thematic content. This works in that participants reason by experience, the schemas contain abstract rules about reasoning. When the participant sees the relationship between the task with which have they been presented and the abstract rules present in

the schema they apply the appropriate rules and thus make the correct selections on the task.

To consider the theory now in more detail the schema that Cheng and Holyoak explained fully in their 1985 paper will be considered here, that is the permission schema (although they also considered obligations in 1989). The permission schema is acquired from those situations where preconditions must be satisfied in order to carry out a particular action. Cheng and Holyoak suggested that this schema consists of four production rules, which can be applied when the conditional in question is of the form: 'If an action is to be taken then the precondition must be satisfied'. The four rules are as follows:

Rule 1: If the action is to be taken then the precondition must be satisfied

Rule 2: If the action is not to be taken then the precondition need not be satisfied

Rule 3: If the precondition is satisfied then the action may be taken

Rule 4: If the precondition is not satisfied then the action must not be taken

In order to understand how these rules facilitate selection of the p and ¬q cards consider the Drinking Age rule that was outlined above. Participants select the ¬q card because rule 4 above tells them that if you are not over 19 then you must not drink alcohol, (where p corresponds to the action and q to the precondition). Thus from rule 4 above participants understand that if the precondition is not satisfied (being 19 years or older), then the action, (drinking alcohol), must not be taken. Participants know from this that they must select the underage drinker for further examination, as the precondition for the action has not been satisfied. The common error of selecting the q

card on the selection task is avoided because participants know from rule 3 that they do not need to check any card where the precondition has been satisfied. Specifically, rules 1 & 4 make determinate prescriptions, whereas rules 2 & 3 make indeterminate prescriptions.

It can be seen from this example that it is possible for the theory to explain how facilitation occurs on thematic versions of the task where the conditional can be seen to fall into the form of a permission. The theory can explain why participants perform well on the Sears task although they have no experience of being a shop manager or assistant. The action is buying goods to the value of $30 or more, and the pre-condition is acquiring the signature of the manager, thus the permission schema is retrieved and applied to the task, and participants accordingly select the p and not-q cards. It is the case that there may need to be some form of authority figure or reason to elicit the permission schema. For example, the policeman in the Drinking Age rule (Pollard and Evans, 1987) or the prior knowledge of the rule in the postal task, where the participants were aware that the rule has been announced by an authority figure. Consequently, they were able to see the rule as a permission based rule, and thus they could elicit the permission schema, and subsequently solve the task correctly, selecting the p and not-q cards.

Cheng and Holyoak (1985) have shown the importance of the ability of participants to see the task in terms of permission, and thus allowing the retrieval of the necessary schema. In order to do so they used the postal task, as described above, but used two groups of participants, some who had experience of such a postal rule (Hong Kong participants) and others who had no experience of such a rule (US participants).

However, the main difference between this version of the task and the original, was that Cheng and Holyoak included a rationale for some participants, stating that the purpose for the rule was to increase profit from personal mail which is nearly always sealed. When the rationale was present the US participants performed as well as the Hong Kong participants despite the fact that the US participants had no real world experience of such a rule before. Cheng and Holyoak concluded that the rise in performance of those participants who had no prior experience of the rule was due to the rationale facilitating the retrieval of the permission schema by participants.

In addition to this Cheng and Holyoak (1985) even managed to produce facilitated performance of an arbitrary task. Again, using a content that could evoke the permission schema, and enable the participants to solve the task correctly did this. Participants were told that they were an authority figure who had the duty of checking regulations of the form "If one is to take action 'A' then one must fulfil precondition 'P'." Participants were then presented with four cards, which showed on one side whether an action had been taken or not, and on the other side whether the precondition had been fulfilled or not. The four cards showed one of each contingency, e.g. action taken, action not taken, precondition satisfied, precondition not satisfied.

Unlike previous arbitrary rules, where performance was very low (around 19%), performance on the task used by Cheng and Holyoak was up to 61%. Finally, with reference to the permission schema Cheng, Holyoak, Nisbett and Oliver (1986) found that if participants had abstract training with schemas then facilitation on the selection task was elicited. Cheng et al. (1986) also introduced schemas for obligations, which are also a major factor in deontic reasoning.

3.4.2 Criticisms of Pragmatic Reasoning Schema Theory

The pragmatic Reasoning Schema theory has been criticised on both methodological and theoretical grounds and these criticisms will be considered here. Jackson and Griggs (1990) criticised Cheng and Holyoak on methodological grounds, because Cheng and Holyoak used explicit negatives on their cards, and Evans (1983) had shown that explicit negatives decrease the extent of matching bias on the selection task. On abstract rules (both permission and obligation) Jackson and Griggs found that facilitation disappears when implicit negatives are used. It is clear from these findings that this does not bode well for PRS, thus shaking one of the major tenets of the theory, because the findings of Cheng and Holyoak may have been dependent on the use of explicit negatives rather than because schemas were being evoked.

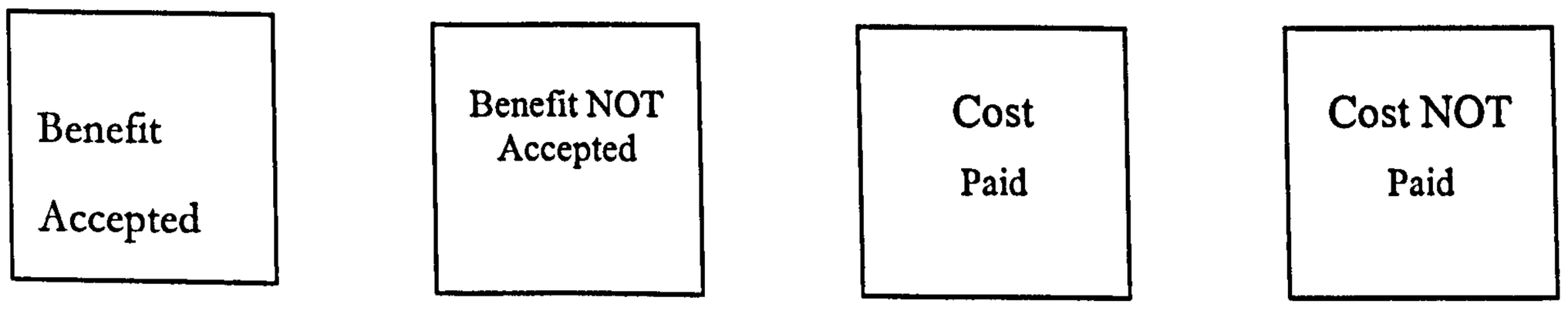
The theory has also been criticised on theoretical grounds. For example Evans (1991) criticised the theory for being too narrow, it is unable to account for the competence and errors that have been observed on abstract reasoning problems, it can only account for thematic versions of the task. O'Brien (1993) also made this criticism of PRS theory - the theory can only account for performance on 'quasi-selection tasks' such as the drinking age version of the task. Also the theory has been criticised because it is not properly formulated, for example, Oaksford and Chater (1993) have criticised PRS theory because it does not explain how beliefs are updated. Given that the theory emphasises the role of domain specific knowledge there should be some acknowledgement of this process. A final critique of the Pragmatic Reasoning Schema Theory comes in the form of an alternative domain-specific theory: Social Contract Theory (SCT). This theory, and the critiques of it are outlined below.

3.4.3 Social Contract Theory

This theory is an evolutionary based approach, in which it is claimed that humans are innately programmed to look for cheaters. Humans must try to maximise the benefit from social situations. Individuals must try to cooperate with each other in order to provide the most mutual benefit from such social contracts. In such social contracts people are obliged to pay costs in order to receive benefits. Cosmides (1989) argued that humans are acutely aware of the potential of others taking the benefits without paying the cost.

Cosmides argued that the "look for cheaters" algorithm would produce logical facilitation on a standard selection task (that is p and not-q). Cosmides (1989) and Cosmides and Tooby (1992) argued that there were two types of social contract, the standard social contract, and the switched social contract. The standard social contract was of the form '*If you take a benefit then you pay the cost*' whereas the switched social contract was of the form '*If you pay the cost then you take the benefit*'. Obviously, rules in this form lend themselves to study via the selection task, the cards used by Cosmides are shown in Figure 3.6.

Figure 3.6 Cards Used by Cosmides (1989)



Cosmides (1989).

Depending on which of the social contracts is being considered, the logical category of the card varies.

Using the standard social contract the cards are as follows: Benefit accepted = p, Benefit not accepted = not-p, Cost paid = q, cost not paid = not-q. However, for the switched social contract the logical categories are Benefit accepted q, Benefit not accepted not-q, Cost paid p, cost not paid not-p.

Importantly, despite the logical category of the cards the participants should still select the 'benefit accepted' and the 'cost not paid' cards. These are the two that represent potential cheaters, and if humans have a 'look for cheaters' algorithm, then these are the cards that participants should select. In addition to this participants should ignore the 'benefit not accepted' and the 'cost paid' cards, as these cards could not represent cheaters. Cosmides argued that any organism which did not have this 'search for cheaters' algorithm would be selected out, hence the evolutionary aspect of her theory. In summary then, while formal logic states that we should always accept the p and not-q cards, Cosmides claimed that this should not be the case in her switched social contract where she suggests that our cheater detection should lead us to accept the not-p and q cards. Cosmides' predictions were upheld in that participants did choose the 'benefit accepted' and 'cost not paid' cards regardless of the logical status of the cards.

Cosmides argued that the facilitation that was shown by Cheng and Holyoak (1985) was due to the rules that they used expressing social contracts. Cosmides backed up this claim by producing versions of the selection task that were not social contracts, but yet

were still permission rules. On tasks such as this there was significantly less facilitation than with social contract rules. This finding is supported by evidence from Gigerenzer and Hug (1992), who found similar results. However, on the non-social contract versions of the selection task performance was above the 10% level of the abstract selection tasks. For the non-social contract based problems, Cosmides and Gigerenzer and Hug reported correct responding of between 30% and 52%. Cosmides also claimed that her theory could account for existing data on the selection task. She claimed that all previous versions of the selection task that had produced reliable facilitation were social contracts (e.g. Johnson-Laird, Legrenzi and Legrenzi, 1972), whereas unreliable, or non-facilitatory versions of the task were not social contracts (e.g. Wason and Shapiro, 1971; Manktelow and Evans, 1979).

3.4.4 Criticisms of Social Contract Theory

Cheng and Holyoak responded to Cosmides' theory by claiming that not all the facilitating contexts can be seen as being social exchanges, they point out the Drinking Age rule and the Sears problem are examples of this. Platt and Griggs (1983) found some support for SCT in that facilitation was not as strong when the cost-benefit structure was removed from problems. They also found that performance was better when the deontic term 'must' was included in the rules, although it could be argued that this also supports PRS theory.

A number of theoretical criticisms have been made of the theory. For example, O'Brien (1993) has criticised SCT, like PRS, on the grounds that the scope of the theory is too narrow, it can only explain performance on what he calls 'quasi selection tasks', such as

the Drinking Age rule version of the task. O'Brien further criticised SCT in that he sees no *a priori* reason why evolution should have provided humans with domain-specific processes, and not to have endowed human thought with some general processes. Indeed he goes on to claim that it would be of greater evolutionary benefit for humans to have general rules rather than domain-specific rules as these would allow for reasoning to still take place even though the environment may change. However, Cosmides responded to this claim by arguing that 2 adaptive rules would always be better than 1 general rule, and that evolution would shape specific adaptive rules, rather than general rules for reasoning.

A further criticism was made by Manktelow and Over (1990): there are instances of the selection task which participants can solve, but that do not rely on a benefit-cost structure in order for facilitation to occur. This was found to be the case using the AIDS rule of Manktelow and Over (1990), this rule stated that "If you clear up spilt blood, you must wear rubber gloves". There is no way that cleaning up spilt blood can be seen as being a benefit, for the cost of having to wear rubber gloves, and yet participants select the p and not-q cards with very little difficulty on this task.

It appears from this version of the task that the primacy of the benefit-cost relationship suggested by Cosmides (1989) is not as reliable as originally claimed. Manktelow and Over (1991) agree with Cosmides in that she introduced the idea of utilities into the field, but that the scope of utilities should be more general than a strict, narrow cost-benefit structure. Participants would *prefer* to wear rubber gloves if they were to clear up spilt blood, as it is in their own best interests. Both Pragmatic Reasoning Schema Theory and Social Contract theory will be considered in more detail in the General

Discussion. The role of utility in the selections which participants make on the selection task will be considered in greater detail in the next chapter.

CHAPTER 4

Page

4.1 Utility in Reasoning

4.1.1 Chapter Overview

The purpose of this chapter is to address closely the role of utility and social roles in reasoning. As discussed in Chapter three Cosmides suggested that utility may be important in reasoning about conditionals. However, Manktelow and Over (1991; 1992) and Over and Manktelow (1993) have argued that Cosmides' cost-benefit structure was too narrow, and that utility should be considered more generally. Thus this chapter will examine closely the empirical work of Manktelow and Over (1991), and the suggestions from this for a revision to the mental models theory. Finally it will be argued that if utility (preferring one outcome to another) plays a role then so should probability (the likelihood of any given outcome). This will clearly show the rationale for the current research programme.

4.2 The Role of Utility in the Selection Task

4.2.1 An Introduction to Utility

It was the research of Manktelow and Over (1991) that first studied the role of social roles and utilities with regard to deontic reasoning, and it has caused a number of revisions or calls for revisions to existing theories (Johnson-Laird and Byrne, 1992; Holyoak and Cheng, 1995a). As was outlined in Chapter three, there has been a lot of research that has shown that there is an effect of content in reasoning. This has posed some problems for theories which claim that reasoning is mediated by content free inference rules (see Evans, 1989;

Manktelow and Over, 1987, 1990; Cheng and Holyoak, 1985; but see also O'Brien, 1993, 1995).

As was mentioned in Chapter 3, Manktelow and Over (1991) considered Cosmides' Social Contract Theory to be too narrow in its benefit-cost structure. Subsequently, they designed a task that did show facilitation in participants' performance, and yet did not fit into the benefit-cost structure which Cosmides claimed was essential for facilitation on the selection task (the AIDS rule, Manktelow and Over, 1990a). Therefore, Manktelow and Over (1991) argued that it is the utility of a particular outcome that determines the choices which humans make. Thus although it was Cosmides, as cited in Manktelow and Over (1990b) who pointed out the importance of utilities in reasoning, Manktelow and Over used it in a more general way. Therefore, both utilities and social structures, which Manktelow and Over argued both play a role in mediating research, will be considered in some detail, as it was the research findings about utility which motivated the research to be presented here.

However, there is an important issue to note: Manktelow and Over clash with Cheng and Holyoak with regard as to what is a conditional permission and what is a conditional obligation. Manktelow and Over claim that what Cheng and Holyoak term as being a conditional permission is in fact, a conditional obligation, because of the emphasis on what one *ought to* do. Due to the importance of pragmatics in deontic reasoning it is imperative that there is a clear understanding of what is a permission and what is an obligation, between researchers in the field.

4.2.2 Empirical Evidence for Social Roles and Utility

Manktelow and Over (1991) studied the role of utilities by using the following statement, given to a son by his mother '*If you tidy your room then you may go out to play*'. They claim that from this statement it is clear that the mother prefers tidy rooms to untidy rooms and thus she places a higher utility on the boy tidying his room than not. Conversely, one can assume that the son places a higher utility on going out to play than not doing so. If this were not the case, then it would be foolish for the mother to have uttered the statement in the first place, as the payoff of going out for tidying the room would be of no value to the son. It is also important to note here that pragmatically it seems to be the case that the only way that the son can get permission to go out is by tidying his room.

Manktelow and Over term the person who utters the rule as the agent, and the person who is the subject of the rule as the actor. Following from this they claim that there are four ways in which the rule can truly be said to be broken. These are as follows:

1. The agent (mother) sees that the son (actor) has tidied his room but does not allow the actor to go out.
2. The agent does not see that the actor has tidied his room, and yet she allows him to go out.
3. The actor (son) tidies his room, and yet he does not go out
4. The actor does not tidy his room, and yet he goes out anyway.

Manktelow and Over looked at cases 1 and 4 above, and they compare these two cases to the benefit-cost structure of Cosmides (1989). Manktelow and Over point out that case

four is similar to the benefit-cost structure, in that the son takes the benefit (going out to play), and yet does not pay the necessary cost according to the rule (tidying his room). However, this is not true with case 1 above, where the mother sees that the son has tidied his room, and yet the mother does not let him out. Clearly, this can still be seen as breaking the rule, in that the agent has not kept her side of the deal.

Manktelow and Over argued that even though case 1 does not fit with the structure that Cosmides sees as necessary for correct choices on the selection task, participants would still be aware of it as being a violation of the rule. When they are cued to see the task from the point of view in which this would be a violation they should select the correct cards accordingly.

Participants were presented with the following four cards. Each was understood to show on one side whether or not the room had been tidied, and on the other side whether or not the boy had gone out to play.

Figure 4.1 Cards Used by Manktelow and Over, 1991

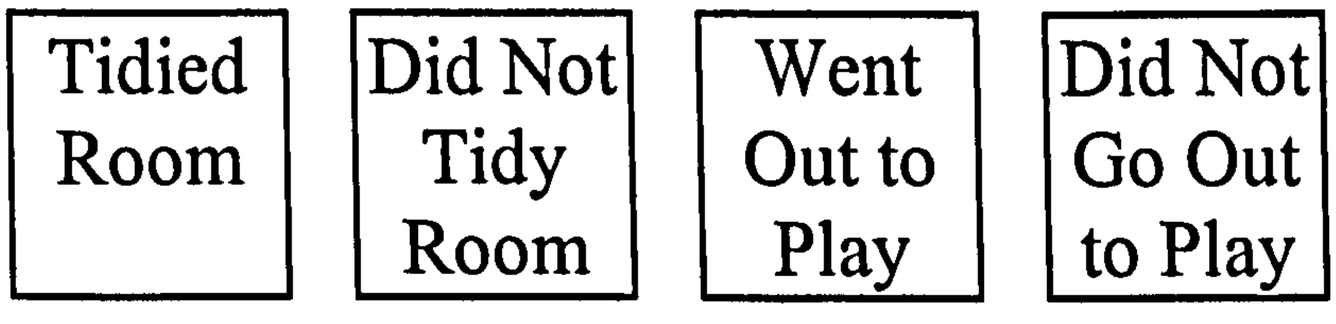


Figure 4.1 Manktelow and Over (1991).

Participants, through a scenario presented with the task, were cued to see the task from either the perspective of the agent (mother) or from the point of view of the actor (son). Clearly, as stated above these roles involve the participants in having different utilities with

regards to the task, and what they see as being violations. Manktelow and Over made the following predictions about the choices that their participants would make on the task. They argued that those participants who were cued to the perspective of the agent (mother) would select the not-p and q cards, and those who were cued to see the task from the perspective of the actor (son) would select the p and not -q cards.

For those participants who were cued to see the task from the perspective of the agent (mother), the predictions were overwhelmingly confirmed, in that the modal selection pattern was to select the not-p and q cards (that is a reversal of the correct selections for the 'normal' selection task). However, this was not the case when the participants were cued to see the task from the perspective of the agent, although the p and not-q cards were selected reasonably frequently, it was not the modal selection pattern.

In order to try to find an explanation for these results Manktelow and Over point out that in the scenario which cued participants to the perspective of the agent, there was a third party who recorded the details on the cards. Therefore, it was felt that a third party should be introduced into the scenario that was designed to cue participants to the perspective of the actor. When this was done the results were that the predicted patterns were found to be the modal selection.

Finally, Manktelow and Over carried out a third experiment in order to test whether the participants understand the deontic context and its utilities and social roles that account for the results. The alternative explanation to this is that participants were relying on personal experience in the above experiments, and so an experiment that used a scenario that differed from a field with which the participants were familiar was designed.

In this third experiment, the rule was "If you spend more than £100, then you may take a free gift". In this case the shop is the agent, and the customer is the actor, and as above there are four ways in which it is possible for the rule to be violated. In this experiment Manktelow and Over made predictions as to what the participants' selections would be for all four cases, as they cued participants to check for each of the four, rather than just for cases 1 and 4 as in the previous experiments. The results followed the predictions: for case 1 the participants selected the p and not-q cards, case 2 the not-p and q cards, case 3 the p and not-q cards, case 4 the not-p and q. Manktelow and Over then claimed that this was evidence that a theory of deontic reasoning must take account of utility and social roles.

4.3. Utility and Mental Models

4.3.1 Revised Mental Models

Manktelow and Over discussed their results in terms of an adapted mental models theory, which took account of utility and social roles. Mental models theory was discussed in depth in Chapter two, and the basics of the theory will not be considered again here, rather the changes to the theory which were called for will be discussed.

Manktelow and Over argued that in addition to representing the possible states of affairs, and the outcomes of actions, mental models should also represent the preferences which people have between these outcomes, hence utilities need to be incorporated. However, Johnson-Laird and Byrne (1992) argued that there was no need to incorporate utilities into mental models theory. Alternatively, they suggested that participants consider only those

cards that are explicitly represented in their models, and thus in order for a participant to have full insight into the task, they must construct fully explicit models.

Johnson-Laird and Byrne (1992) offer their explanation of Manktelow and Over's results in terms of the third experiment that Manktelow and Over used. That is the rule "If you spend more than £100, then you may take a free gift". The participants had four cards in front of them, showing a receipt, and whether or not the customer had taken the free gift or not. Participants were asked to choose those cards that would show whether the rule had been violated or not; obviously this could be a violation by the shop, or by the customer. Johnson-Laird and Byrne considered the rule in terms of '*If p then q*', and the four cards represent: the shopper spent more than £100 (p), the shopper did not spend more than £100 (not-p), the shopper took the free gift (q), and the shopper did not take the free gift (not-q).

Johnson-Laird and Byrne (1992) began their account by outlining the model theory of the selection task (Johnson-Laird and Byrne, 1991) as a result this is where the current account will begin. Using a conditional such as "If there is a letter A, then there is a number 2". One explicit model, and one implicit model, as discussed in Chapter 2 would then represent this:

[A] 2

...

Participants presented with the task only consider those cards that have been explicitly represented in the models, the theory states. As a result of this the participants select the A and 2 cards, and as was discussed in Chapter 3, this is the modal choice when the task involves a indicative conditional of the form above.

However, it is possible for the models to be fleshed out into three explicit models, which each represent a possible state of affairs, these three models would be:

A	2
¬ A	2
¬ A	¬ 2

When the models are fully fleshed out like this the participants now consider the ¬A and ¬2 cards, which they would not have done in the first model above. Participants do not select the ¬A card because this could occur with either the 2 or ¬2 cards, and thus this card is not able to determine whether the rule is true or not. However, the ¬2 card could determine the rule as false if it occurred with the A card, and so the participants add it to their selections.

Therefore, the model theory proposes that any context or content that causes the participants to flesh out the models will result in them selecting the not-q card. Such cases of facilitation were discussed earlier. Johnson-Laird and Byrne suggest that the search for violations will lead to participants fleshing out their models fully, and hence performance is facilitated in a deontic selection task. However, they go on to contend that there are versions of the selection task which lead to a fleshing out of models, but yet have nothing to do with deontic contexts (Oakhill and Johnson-Laird, 1985). Thus Johnson-Laird and Byrne conclude that no account of performance on the selection task can be complete if it relies solely on deontic considerations, as is the case with both Cheng and Holyoak (1985), and Cosmides (1989).

Johnson-Laird and Byrne use the example from Manktelow and Over (1991) when they consider how to apply mental models theory to deontic conditionals, that is: '*If you spend more than £100 then you may take a free gift*'. They then go on to show the models which participants would need to flesh out if they were presented with this conditional in terms of the selection task:

£100 gift

£100 ¬gift

¬£100 ¬gift

All of these represent the permissible situations, that is those situations which do not represent a violation of the rule that the participants have been given, with regard to the conditional, as opposed to the truth/falsity that one would normally consider, whereas, the following model would be impermissible:

¬£100 gift

As this represents the case where a customer spends less than £100, but takes a free gift anyway.

However, as was seen above it is possible that participants interpret a conditional as being a biconditional, and then the permissible and impermissible models are altered. Thus:

£100 gift

¬£100 ¬gift

Are now the permissible situations, whereas the models that represent the impermissible situations are as follows:

£100 ¬gift

¬£100 gift

The first model here could arise by the customer not selecting the gift even though they had fulfilled the necessary requirements, or because the store had gone back on its promise to give gifts to those customers who spent more than £100. The second impermissible model above could come about by the customer taking a free gift that they were not entitled to, or by the shop relaxing the rule and allowing those customers who have not fulfilled the requirements to claim a free gift. Hence Johnson-Laird and Byrne conclude that the results observed by Manktelow and Over come about via the context of the selection task leading the participants to focus on one set of models rather than another.

It was discussed above that Manktelow and Over did not find the predictions that they had expected in one half of their first experiment. When participants were asked to imagine that they were a son who was checking whether his mother had broken the rule, *'If you tidy your room, then you may go out to play'*, it was found that it was necessary to introduce a third party. The effect of this third party was that participants selected the cards which Manktelow and Over predicted that they would (p and not-q cards). With this conditional the models for the impermissible situations are as follows:

tidy ¬play

¬tidy play

It is possible then to see from this that if participants select all four cards then they can check for both types of violation, and this was the modal response found by Manktelow and Over (1991). Johnson-Laird and Byrne claimed that Manktelow and Over changed the instructions to such an extent that it was now made clear to participants that only the serious violation was at stake, that is that the son should tidy his room, and yet not be allowed out to play. Therefore, Johnson-Laird and Byrne claimed that it was not the introduction of a third party that led to the p and not-q cards being selected. Rather, the instructions now pointed the way for the participants to regard the mother seeing the room was not tidy, but still allowing the son out to play as being irrelevant to the task as such.

In their conclusion, Johnson-Laird and Byrne argued that the content of a task can lead to participants constructing an explicit set of models, and yet application of everyday knowledge can lead to the participants interpreting the conditional as a biconditional. Hence they have two impermissible situations in their models, and the context of the task can lead to participants concentrating on one of these two impermissible situations. Thus Johnson-Laird and Byrne claimed that their explanation of the results found by Manktelow and Over (1991), have the advantage of parsimony over that of Manktelow and Over.

However, Manktelow and Over (1992) replied to Johnson-Laird and Byrne. They claimed that when Johnson-Laird and Byrne suggested that there is a serious violation for the son, they are in effect talking about the subjective utility of the violation for the son. Thus Manktelow and Over claimed that Johnson-Laird and Byrne are admitting that participants do consider the utilities in the task, and that they *do* affect the reasoning in which they subsequently engage.

Manktelow and Over (1992) produced new data, which they claimed corroborated their suggestion for the need to consider utilities within a mental models framework in order to account for deontic reasoning. In this experiment subjective utility was varied, so that there were two scenarios with what were termed strong and weak preferences. There was also an experiment that involved participants reasoning using an epistemic conditional, as Johnson-Laird and Byrne (1992) draw comparisons and make similar predictions between epistemic and deontic conditionals.

In the epistemic task there is no real world content, whereas in the in the weak preference task there is no evidence for the participants to judge as to whether one outcome is preferred to another. Finally, in the strong preference task, there is some evidence to show that there may be cheaters who would try to bring one outcome about in preference to another (that is that they may be able to gain a prize if they cheat).

The task is presented in the traditional selection card format, and Manktelow and Over stated that Johnson-Laird and Byrne's account would predict that the modal response for all cases should be not-p and q. Whereas Manktelow and Over predict that this would only be the case in the strong preference condition.

The results were that the modal response was to select only the q card in the epistemic task, although there was no significant difference in the choices that participants made. In the weak preference task both p and p, not-p were chosen 3 times each, and again there was no significant difference between the selections made by participants. Finally, in the strong preference task the modal choice was not-p and q as predicted by Manktelow and Over, and there was a significant difference in the choices made by participants (only four participants made a choice other than not-p, q and this was to select the q card alone).

Therefore, Manktelow and Over claimed that there must be some consideration of utility in deontic reasoning accounts, otherwise, what is the point in reasoning about actions or outcomes if people have no preference between the outcomes?

It should be noted here that Johnson-Laird and Byrne (1995) have revised their account of perspective effects in deontic reasoning, in a version that involves biconditional interpretation, focussing and logical negation. Using Manktelow and Over's conditional:

If you tidy your room then you can go out to play

Johnson-Laird and Byrne argue that this can be given the biconditional interpretation:

If you do not tidy your room then you can not go out to play

Fleshing out the models for these yields:

t	p
¬t	¬p

where t is tidying your room, and p going out to play. The mental models theory relies on the search for counterexamples, there are two possible counterexamples here:

t	¬p
¬t	p

A full test of the conditional would result in the selection of all four cards - as is the case with participants who adopt a neutral perspective (Politzer & Nguyen-Xuan, 1992).

Johnson-Laird and Byrne argued that the perspective adopted by a participant would effect the models constructed. The son will construct the following model:

t p

...

through the application of negation the following counterexample is derived:

t ¬p

The participants who adopt the perspective of the mother represent the following model:

¬t ¬p

again, via negation the following counterexample is yielded

¬t p

In the cases above the participants select the cases that correspond to the counterexamples yielded by the application of negation to their initial models.

Manktelow and Over claimed that it was clear that utilities play a role in mediating reasoning, and thus there is a link here between reasoning research and decision making theory, as subjective utility is traditionally a concept of the latter. Johnson-Laird and Shafir (1994) discussed the increasing links between decision making and reasoning research, which are traditionally two disparate fields of research (see also Doherty and Evans, 1996 on the relationship between reasoning and decision-making). Thus in the tradition of subjective expected utility theory there is scope here for the study of the role of probability in reasoning research, as Manktelow and Over called for. The role of utility, or more generally preferences, has already been considered above, and this has come from decision theory. Decision theory states that not only are preferences important, but also probabilities. A brief, and somewhat crude explanation of the theory of maximising subjective expected utility will now be presented. Given two possibilities to choose from A

or B the following should become a consideration: what preference, if any, is there for A over B or B over A? Manktelow and Over have demonstrated that utility is important in deontic reasoning, but how does probability apply?

Once preferences have been assigned to the possible outcomes, it is necessary to consider now the *likelihood* of each of these outcomes given the possible actions. It is possible to assign a numerical value to each of these factors (utility and probability). It is important to note that these numbers do not really signify anything in the world - just the value of our preference of A over B, and our belief in the likelihood of A given a certain set of actions. It is then possible to calculate subjective expected utility using the following equation:

$$SEU = \sum_{i} s_i U_i$$

where s_i is the subjective probability of the *i*th outcome, and U*i* represents its subjective utility, and *i* ranges over a finite set of mutually exclusive and exhaustive outcomes. According to the theory we should select the action that has the highest possible value following this calculation. It is clear from this brief summary that probability as well as utility has a role in decision making. Given the findings for the role of preferences in reasoning, it is a sound move to consider the role of probability in reasoning.

Some recent work on probability in reasoning will briefly be discussed here before going on to outline the research conducted as part of the current programme. Green, Over and Pyne (1997) investigated the role of probability in reasoning, using the selection task. In their experiments participants were required to estimate the probability of finding a counter example to the conditional that they had been given. Green et al. then explored the

relationship between these estimates and selections on the task. In the two experiments reported Green et al. found that probability estimates were a good predictor of selections made. Following Green (1995) Green et al. (1997) employed externalization procedures that require that the participants think about counter-examples. There were two externalization procedures used. Firstly, participants must try to envisage the different possible alternatives on each card, and secondly, they must identify those cards that could contain potential counter-examples. They argued that results should be considered in terms of subjective expected utility. In combination with the work of Manktelow and Over (1991) there is evidence to suggest that both utility and probability have a role in mediating the reasoning process.

Initial work then would suggest that utility has a role to play in mediating reasoning performance and Green et al. (1997) and Kirby (1994) have produced some evidence to suggest a similar role for probability. It should be noted here that the research was conducted prior to the publication of Green et al.'s paper. The empirical work to examine the role of probability in reasoning will now be presented in chapter 5.

CHAPTER 5

5.1 The Horse Race Tasks

5.1.1 Probability and Reasoning with Conditionals

As was discussed in Chapter 4 Manktelow and Over (1991, 1992) argued that there should be some element of the utilities which participants ascribe to certain outcomes, and that utilities should be clearly represented in a new formulation of mental models theory for explaining deontic reasoning. There appears to be strong evidence, as discussed earlier (Chapter 4), that utilities do play a role in mediating deontic reasoning.

Manktelow and Over (1992) pointed out that it was Cosmides (1989) who first drew attention to the role of utilities in reasoning, based around her notion of a benefit-cost structure. Manktelow and Over, however, claimed that this structure was too limited, and that there should be a more general form of preference represented in a full theory of reasoning, and they termed this utility, which is considered to be a traditional construct of decision making research. Manktelow and Over presented evidence in support of the presence of utility in reasoning. Another construct usually associated with decision making should be considered in relation to deontic reasoning: probability. More specifically, not only should reasoners consider what use a particular outcome is to them (utility), they should also consider how likely the possible outcomes are. Reasoning should be based upon a combination of these factors, as argued in the theory of subjective expected utility, within the decision-making literature (see von Winterfeldt and Edwards, 1986 and Evans and Over, 1996a & b).

It was the aim of the first experiment reported in this thesis to begin to examine the role of probability in reasoning. This was done by designing tasks based around horse racing as this was considered to be a field in which probability is intrinsically involved. It also has the advantage of being an area with which the majority of participants would be familiar, thus avoiding any complications around not understanding the task. It is worthy of note here that there are very few real-world tasks in which odds are stated as explicitly as they are in horse racing. However, there would be a number of manipulations that would emphasise, or not, the importance of winning to the participants, thereby combining both probability and utility in the tasks. Indeed, in practice these two constructs are inextricably linked as participants are likely to have a preference for various outcomes, or a preference of differing degrees for the same outcome.

The three scenarios used here were as follows: one would suggest to the participants that the bet was merely for fun, and winning was not of great importance. The second would stress the importance of avoiding a loss for the participants, and that it was a serious bet that they must win. The final scenario suggested to the participants that winning the bet would be good for them, and yet it was seen as a long shot bet, and not of huge importance. The scenarios will be given here in full.

5.2 Experiment One.

Method

5.2.1 Participants.

The participants in this experiment were seventy-four first year psychology undergraduates at the University of Wolverhampton. All participants took part on an unpaid volunteer basis, and had no prior experience of the selection task.

5.2.2 Materials.

Scenario one: Trivial Bet:

This scenario was designed so that participants would not place a great emphasis on winning the bet, the scenario was presented to participants as follows:

Trivial bet:

You have recently inherited £100,000 from an old relative, one of the less serious things you have decided to do with this unexpected windfall is to place a bet of £100 on a horse. A friend of yours hears of your plans and tells you about the tipster in "Racing Weekly". She tells you that: **If Racing Weekly tips a horse then it wins.**

You know very little about horse racing and as a result of this you think that studying the likelihood of success in your bet will be part of the fun of spending your newly acquired money. You therefore decide to read Racing Weekly and see what the tipster recommends you do.

Your friend has given you four slips of paper on which she has recorded details of horses she backed in four different races. Each slip says whether the horse was tipped by Racing Weekly or not on one side, and on the other side whether the horse won or not.

Which of the slips would you need to turn over in order to judge whether your friend's statement will be likely to lead you to winning your bet? Please tick the appropriate answer box(es).

It is clear from the scenario above that the winning the bet is of little importance, hence the line "one of the less serious things...". The participants were then presented with four cards to choose from. These were Tipped by Racing Weekly (p), Not Tipped by Racing Weekly (not-p), Won (q), Did Not Win (not-q), given that the conditional was of the form if p then q.

Scenario two: Third Party Loss:

This scenario was designed to emphasise the importance of not losing the bet, and this was made more salient due to the fact that the money used was somebody else's.

Third Party Loss;

Your father has given you a large amount of money and has asked you to place a bet on a horse for him. He has left it up to you to decide which horse in which race you will bet on.

You are not sure which horse to back, however, and your knowledge of horse racing is far more limited than you had led your father to believe. Therefore you do not wish to lose your father's money, because he will be greatly annoyed that you have let him down, and you yourself will look extremely foolish for having lied to your father in the first place. Fortunately a friend tells you about the tipster in "Racing Weekly": She states that **If Racing Weekly tips a horse then it wins.**

Obviously you want to know how confident you can be in your friend's statement, so that you know whether to risk your father's money, and your pride, on Racing Weekly's next tip. Your friend has given you four slips on which she has recorded details of horses she backed in four different races. Each slip says whether the horse was tipped by Racing Weekly or not on one side, and on the other side whether the horse won or not.

Which of the slips would you need to turn over to judge whether your friend's statement will be likely to lead to you not losing your father's bet, and

therefore avoid your father finding out that you have deceived him? Please tick the appropriate answer box(es).

Therefore, as was mentioned above it is important to the subject that they win this bet, not only for themselves, but also for their father, hence the introduction of the third party, in that the participant must choose carefully, as their choices do not only affect them. Again the participants were presented with the same four cards as with the other scenarios.

Finally, the last scenario which the participants were presented, was the Low Utility scenario, in which participants would like to win, but yet the consequences are not dire if they fail to do so, that is, the participants would benefit from a win, but losing is not terribly serious.

Scenario Three: Low Utility:

This scenario outlines the bet as being a long shot, and not of the greatest importance, and thus the participants are not massively cued to think about the choices that they make.

Low Utility:

Your friends are planning to go on a camping holiday in Germany during the upcoming summer vacation, and they have asked if you would like to go with them.

You get on reasonably well with all those who are planning to go away and would quite like to go along.

All your friends have jobs and can easily afford the travel costs involved. However, your grant does not really allow you to pay for yourself to go abroad.

As a long shot to afford to be able to go on the trip you decide to place a small bet on a horse in order to pay for your travel. However you do not know a great deal about horse racing, yet a friend tells you about the tipster in "Racing Weekly". She states: **If Racing Weekly tips a horse then it wins.**

Your friend has given you four slips of paper on which she has recorded details of horses she backed in four different races. Each slip says whether the horse was tipped by Racing Weekly or not on one side, and on the other side whether the horse won or not.

Which of the following slips would you need to turn over to judge whether your friend's statement will be likely to lead you to winning your bet? Please tick the appropriate answer box(es).

Again the participants were presented with the same cards as they were with the other scenarios above. An example of the response sheet can be found in appendix 1.

5.2.3 Design.

The conditional with which the participants were presented in this task was an indicative conditional, rather than a deontic conditional, thus participants were expected to test the truth or falsity of the rule. In this task what the participants are trying to do is to assess the likelihood of q given p, which is how likely is it that the horse will win given that it has been tipped? Therefore, in terms of predictions regarding the selections of the participants in this experiment, they would be expected to select the p,q and not-q cards. These would be the rational selections because if there is a not-q on the reverse of the p card, then this would suggest unreliability on the part of the tipster, whereas, a q would suggest that the tipster is reliable. The participants should hope to find a p on the reverse of the q card, indicating that the tipster has tipped any horse that wins, although a not-q would not necessarily suggest unreliability.

Finally, the not-q card, which, as discussed above, is the card that participants often ignore on indicative tasks. The participants should select this card because if they find a p on the other side then this would suggest unreliability on the part of the tipster, as this would mean that horses that lost their races had been tipped by the Racing Weekly. Therefore, as the importance of winning was increased in the scenario, then it was predicted that the selections of the not-q card would increase, as participants do not usually select this card. It is predicted that the participants would select increasingly more not-q cards from the Trivial Bet scenario through the Low Utility scenario, to the Third Party Loss scenario. The not-p (was not tipped) is of no use to the participants as this does not help them to assess the reliability of the tipster which is what the task requires them to do. In summary then,

the predicted pattern of selection is p, q not-q. The not-q card selections should rise because participants are trying to assess the likelihood of q given p.

5.2.4 Procedure

The cards were presented in two different orders, and the different scenarios were handed to the participants randomly during a lecture. The participants were instructed not to confer with each other, and to take all the time they needed to complete the task. 3 task types and two order books, giving six answer books in all. An example of a response sheet can be found in appendix 1.

5.3 Results

Table 5.1 Results for Trivial Bet Scenario

The results for Experiment 1a can be seen in the tables below:

Percentage Frequencies of Card Selections in the Trivial Bet Scenario

Tipped (p)	Not Tipped (not-p)	Won (q)	Did Not Win (not-q)
26	9	21	12
93%	32%	75%	43%

Combinations of Selections

p, q not-q	q	p	p, q	p, not-q	p, not-p	All Cards
3	2	5	10	3	1	6
10%	7%	17%	33%	10%	3%	20%

N = 28

Table 5.2 Results for Third Part Loss Scenario

The results for Experiment 1b can be seen in the tables below:

Percentage Frequencies of Card Selections in the Third Party Loss Scenario

Tipped (p)	Not Tipped (not-p)	Won (q)	Did Not Win (not-q)
15	5	15	10
79%	26%	79%	53%

Combinations of Selections

p, q not-q	q	p	p, q	Not-q	All Cards
4	3	3	3	1	5
21%	16%	16%	16%	5%	26%

N = 19

Table 5.3 Results for Low Utility Scenario

The results for Experiment 1c can be seen in the tables below:

Percentage Frequencies of Card Selections in the Low Utility Scenario

Tipped (p)	Not Tipped (not-p)	Won (q)	Did Not Win (not-q)
24	6	18	7
89%	22%	67%	26%

Combinations of selections

p, q not-q	q	p	p, q	p, not-q	p, not-p	All Cards
2	4	5	9	2	2	3
7%	15%	19%	33%	7%	7%	11%

N = 27

Analysis of variance revealed no significant differences between the conditions for all four cards. An index was also computed based on Pollards Indices (Pollard and Evans, 1987). However, due to the nature of the prediction made (that participants would select the p, q and not-q cards) this procedure needed to be modified slightly. Originally participants would be given 1 point for every logically correct selection made, and one point would be subtracted for every logically incorrect selection made, thus producing an overall score with a range form –2 to +2. However, in the present experiment participants were given 1 point for every predicted card selected (p, q not-q), and one point was subtracted for every non-predicted selection (not-p). This procedure will result in the loss of some data, but does allow for cross-condition comparison for the predicted selection pattern. Thus this computation will be referred to as the *rational* index, and will range from -1 to +3. Matching indices were also produced using the following equation (p + q)-(not-p + not-q). These indices give a score within the range –2 to +2. There were no significant differences between the three conditions on either of these indices. A graph showing the mean rational indices by group is given in fig. 5.1, and a graph showing the mean matching indices by group is given in fig. 5.2.

Figure 5.1 Bar Chart Showing Mean Rational Indices for Experiments 1a-1c

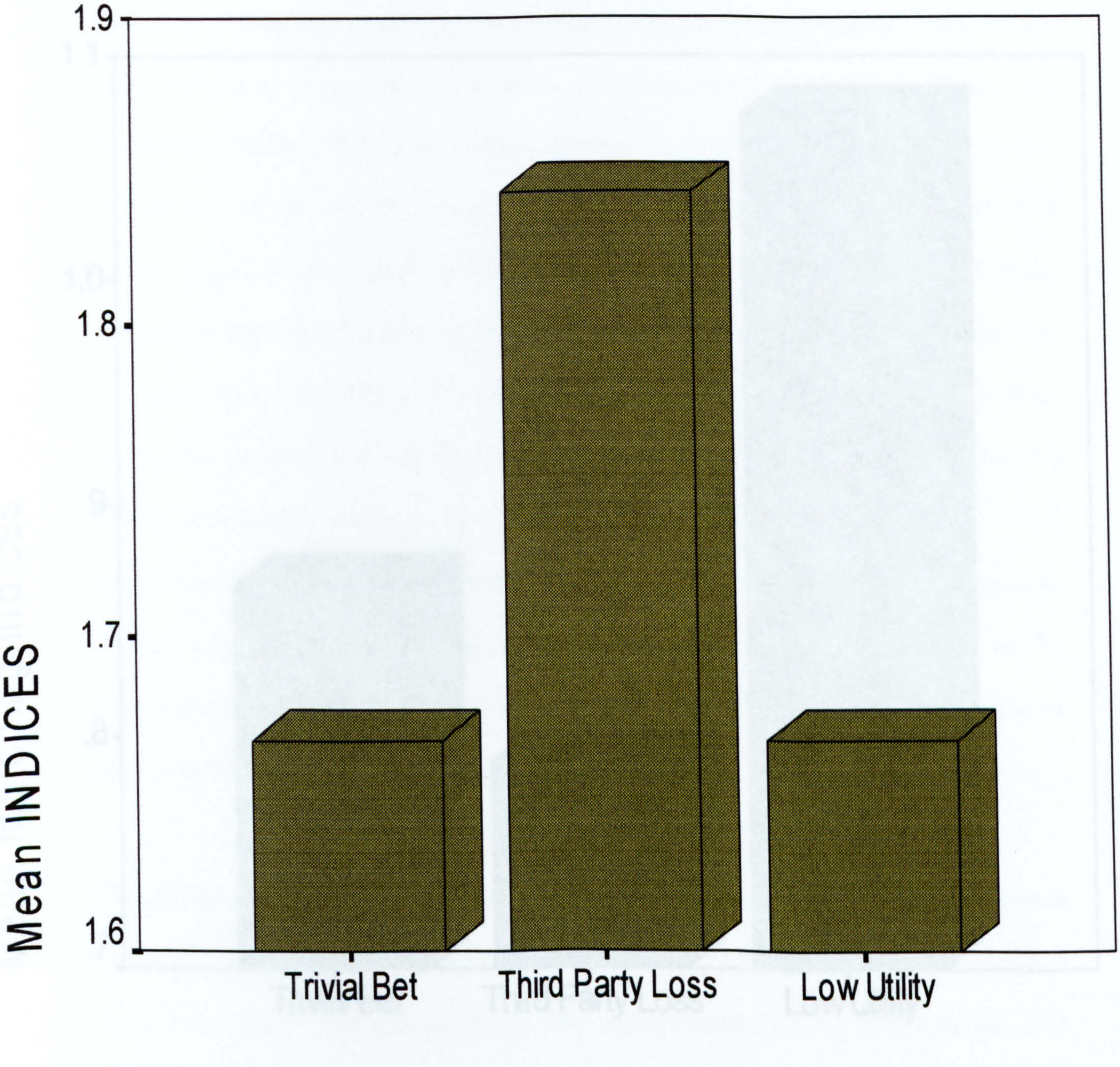
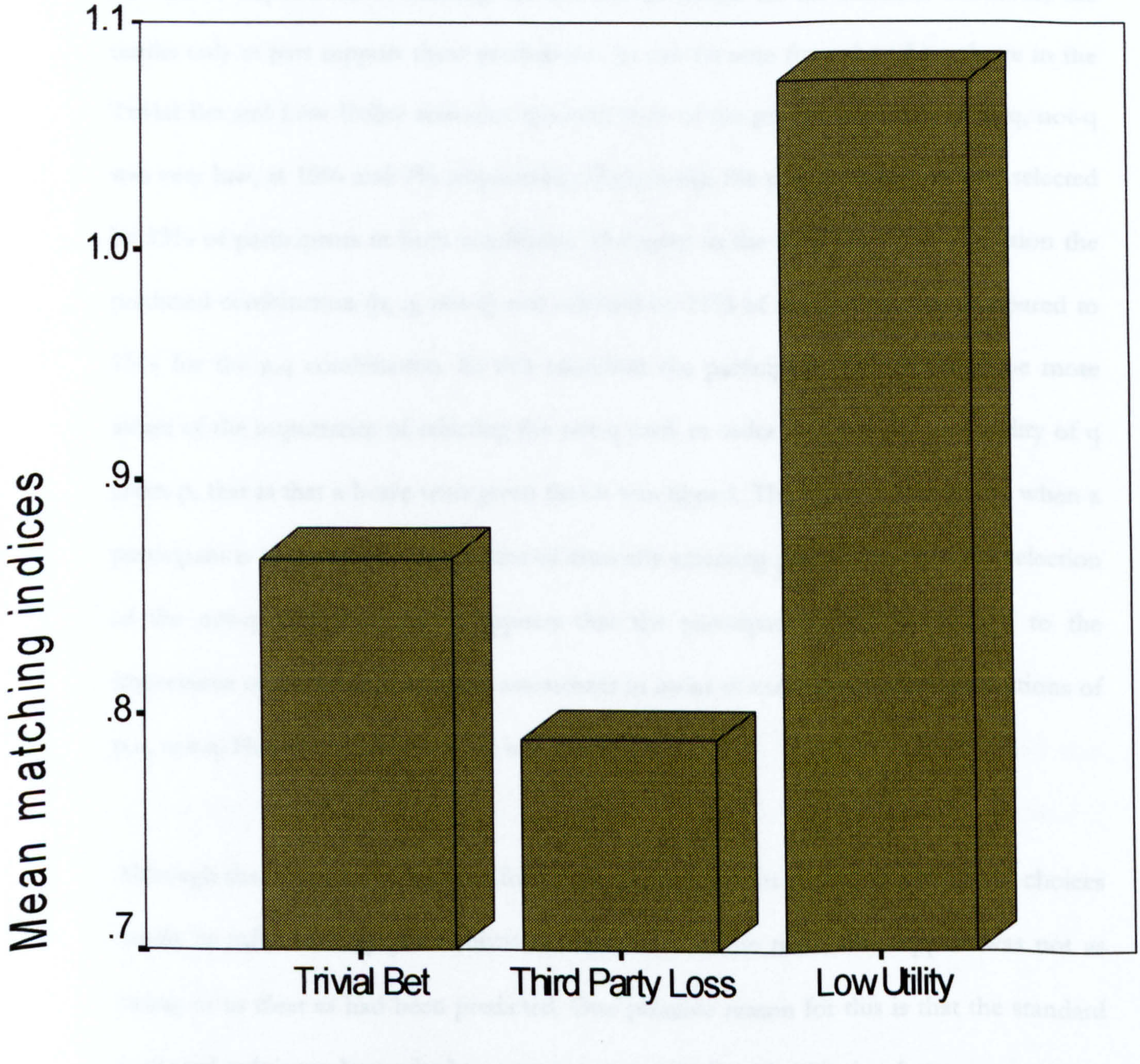


Figure 5.2 Bar Chart Showing Mean Rational Indices for Experiments 1a-1c



5.4 Discussion

It was predicted that the participants would increasingly the select the not-q card as they were trying to assess the likelihood of q given p, and also that these selections would increase as importance of winning the bet was increased via the scenario. However, the results only in part support these predictions. As can be seen from the tables above in the Trivial Bet and Low Utility scenarios the selections of the predicted pattern of p, q, not-q was very low, at 10% and 7% respectively. Conversely, the p, q combination was selected by 33% of participants in both conditions. However, in the third party loss condition the predicted combination (p, q, not-q) was selected by 21% of the participants, compared to 15% for the p,q combination. In this condition the participants did appear to be more aware of the importance of selecting the not-q card, in order to assess the probability of q given p, that is that a horse wins given that it was tipped. Therefore, it seems that when a participant is aware of the importance of correctly assessing probability, then the selection of the not-q card increases. it appears that the participant needs to be cued to the importance of a correct probability assessment in order to make the predicted selections of p,q, not-q. However, this difference was not significant.

Although there was some support found for the predictions that the participants' choices would be influenced by probabilistic manipulations in the tasks, this support was not as strong or as clear as had been predicted. One possible reason for this is that the standard four-card task is not best suited to experiments in which probability is a factor.

Another factor to be considered when testing participants on tasks that involve probability was raised by Gigerenzer (1991). Gigerenzer suggests that the mind is an intuitive

frequentist: when humans make probability judgments they do not base them on a single case, rather they examine occurrences over time, and then they make a judgment based on this. Therefore, the second set of experiments was based on this idea, and gave participants the opportunity to measure the tipster over time in order for them to make a sound probability judgment about the tipster. In order top do this the Large Array Selection Task (LAST) was designed, as it was thought that this would give a more sensitive measure of the selections which the participants make with regard to probability judgments. Briefly, the LAST is a version of the selection where all versions of the cards are presented a number of times. The scenarios used, and the LAST itself are discussed in more detail below, in the second part of this chapter.

5.5 The Frequentist Approach.

5.5.1 Are Humans Frequentists?

Experiment 1 suggested that, participants would select the cards that were relevant to making a sound probability judgment when they are given enough incentive to do so by the content of the task. However, it has been argued that humans may be by nature frequentists, rather than Bayesians (Gigerenzer, 1991, 1996), and, if so, it is important that tasks are presented to participants in this form, so as to get the clearest measure of participants' reasoning. It is therefore important to consider exactly what this frequentist approach involves, and how the approach could be applied to the selection task.

Gigerenzer suggested that the mind might be an intuitive frequentist (see also Fiedler, 1988 and Tversky and Kahneman, 1989). He interpreted classic tasks into a frequentist framework, and found that cognitive illusions, that had previously been considered to be robust phenomena, could be made to disappear. He argued that humans make far better

probability judgments when the tasks are presented to them in such a framework. Consider the following tasks which Gigerenzer presented to his participants.

The Linda problem is a well-known problem that was presented to participants by Tversky and Kahneman (1983), in this problem participants are given a description of a person and are required to make a decision about that person.

> Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in antinuclear demonstrations.

Participants were then asked which of the two following alternatives was more probable:

Linda is a bank teller (T)

Linda is a bank teller and is active in the feminist movement (T&F)

When Tversky and Kahneman presented this to their participants they found that 85% of their participants chose the T&F option. However, Tversky and Kahneman claimed that this was incorrect, that they should choose T because the probability of a conjunction cannot possibly be greater than one of its constituents. For this result Tversky and Kahneman offer the representativeness heuristic (a heuristic being a short cut to an answer which is usually successful, but not always, unlike an algorithm). By employing this heuristic participants base their choice on the match between the description of Linda, and the two alternatives with which the participants are presented. As participants think that Linda is

described as being representative of a feminist in the description they select T&F as being more probable than just T.

However, Gigerenzer goes on to suggest that there is fallacy here. Those participants who select T&F are not violating probability theory. This is where Gigerenzer introduces the concept of frequentist theory (Gigerenzer, 1991). To a frequentist this problem has nothing to do with probability theory. Participants are asked for the probability of a single event, not for frequencies. Indeed Gigerenzer quotes the statistician Barnard who claimed that such judgments should be treated in the context of psychoanalysis, not probability theory. (Although as Gigerenzer notes, he should have said cognitive or social psychology, however, the point he makes remains). Gigerenzer concedes that the conjunction fallacy above is a violation of some subjective theories of probability, such as Bayesian theory. However, it is not a violation of the major view of probability, the frequentist conception.

Fiedler (1988) presented a version of the Linda problem in a frequentist framework, and suggested that this would lead to very different results from those reported by Tversky and Kahneman as above. He suggested that this would be the case if the mind was a frequentist, and thus very sensitive to the difference between single cases and frequencies. Gigerenzer cites Fiedler (1988) as an example of the Linda problem framed in a frequentist framework. Participants were presented with the description of Linda as above, and then asked:

There are 100 persons who fit the description above (i.e. Linda's). How many of them are:

(a) bank tellers

(b) bank tellers and active in the feminist movement.

It was found that when the problem was phrased in a frequentist framework such as this, the conjunction fallacy, as described above, almost disappears. In Fiedler's frequentist versions of the task the conjunction fallacy dropped as low as 17%, compared to 85% in Tversky and Kahneman's original study, and 91% in Fiedler's replication of Tversky and Kahneman's experiment. Therefore, Gigerenzer (1991) argued that humans approach such tasks with frequentist framework in place in their minds, and thus when tasks are presented in this framework, participants perform better than when tasks are presented in a 'single shot' framework.

Thus, in the light of Gigerenzer's arguments the selection task was developed into a frequentist framework, so as to decrease any chance of the participants being influenced by the form of the task. It also allows for a more sensitive measure of participants' choices, as the aim was now to measure the role of probability in conditional reasoning. For example, in the single shot version of the task the participants may not consider the not-q card to be important enough to select. However, if presented with 5 not-q cards as in the LAST the participants may select 2 of them, thus the measure is more sensitive. Again scenarios to influence the importance of winning were presented to participants, one scenario where the participants must try to avoid losing the bet, and one where the participants must try to win the bet. Although initially these may appear to be the same thing, there is evidence to suggest that participants are more sensitive to losses rather than to gains (Kahneman and Tversky, 1979). However the scenarios were now followed by the Large Array Selection Task (LAST - see Appendix 2). Thus it is predicted that those participants who where trying to avoid loss would show a higher level of selection of the not-q card than those

trying to win the bet. A further methodological change that was made here was to add a fifth card to the four that were used in experiment one. This was a second not-q card, NO TIP GIVEN; this card was similar to the fifth card added to the task by Johnson-Laird, Legrenzi and Legrenzi (1972). However, Johnson-Laird et al. predicted that this card would be selected whereas here it is predicted that this card will not be chosen. The reason for this prediction was that the card does not help the reasoner to decide on the likelihood of q given p. The card was added to the task in order to test the prediction that it would not be chosen.

5.6 Experiment 2: The Large Array Selection Task

Method

5.6.1 Participants.

The participants in this experiment were 37 first year biomedical students enrolled at the University of Wolverhampton stood as participants in this experiment. None of them had any prior experience of the selection task.

5.6.2 Materials.

There were two scenarios here that were as follows:

Experiment. 2a. Scenario One: Frequentist Loss:

You have decided to place a large bet on a horse race. However, you're not sure which horse to back, and you want to give yourself the lowest chance of losing all your money. A friend tells you about the tipster in "Racing Weekly": She states that ***"If Racing Weekly tips a horse then it wins".***

You want to see how confident you can be in this statement, so that you know whether to risk Racing Weekly's next tip. Your friend has given you copies of her betting slips from the past month on which she has recorded details of all the horses she placed bets on in different races. Each slip says on one side whether the horse was tipped by Racing Weekly or not, and on the other side whether it won or not.

Which of the slips would you need to turn over to judge whether your friend's statement will be likely to lead to you not losing your bet? Please tick the appropriate answer box(es).

Please turn over the page.

Expt. 2b. Scenario Two: Frequentist Win:

You have decided to place a bet on a horse race. However, you're not sure which horse to back, and you want to give yourself the greatest chance of winning some money. A friend tells you about the tipster in "Racing Weekly": She states that ***"If Racing Weekly tips a horse then it wins".***

You want to see how confident you can be in this statement, so that you know whether to follow Racing Weekly's next tip. Your friend has given you copies of her betting slips from the past month on which she has recorded details of all the horses she placed bets on in different races. Each slip says on one side whether the horse was tipped by Racing Weekly or not, and on the other side whether it won or not.

Which of the slips would you need to turn over to judge whether your friend's statement will be likely to lead to you winning your bet? Please tick the appropriate answer box(es).

Please turn over the page.

There were three randomised card orders, making six booklets in all (that is two scenarios, and three card orders).

5.6.3 Design.

It can be seen that there is a very subtle difference between these two scenarios: one emphasises that the participants should try to avoid losing their money, while the other emphasises the chance of winning the bet, and therefore, making a gain. As was mentioned earlier there is evidence to show that participants are more aware of avoiding loss rather than making again for themselves. The cards presented to the participants were in a frequentist format, for the reasons outlined above (the response sheet can be found in appendix 2).

Again, the prediction is that the participants would select the p, q and not-q cards, for the same reasons as were outlined earlier. However, it is suggested that the participants who are given the scenario in which winning is emphasised would be less likely to select the not-q card than those who are given the scenario in which the subject is reminded that they should try to avoid losing. 2 scenarios and two randomised card orders, making four booklets in each.

5.6.4 Procedure.

The experiment was administered to the participants during their normal classes, and they were randomly given one of the two scenarios. Participants were asked not to confer and the experimenter remained with the participants to ensure that this was the case. The participants were given as long as they needed to complete the task, although most finished within ten minutes. Finally, participants were thanked for their participation, and told of the broad aims of the experiment.

5.7 Results

The percentage frequencies of card selections made by participants in this experiment can be seen in the tables below: (The percentages are calculated by taking the number of each card type selected divided by the total number of those type of cards available for selection and then multiplying by 100).

The results from these experiments are shown overleaf in Tables 5.4, and 5.5.

Table 5.4 Results for The Frequentist Loss Scenario

The results for Experiment 2a can be seen below:

Percentage Frequencies of Card Selections in the Frequentist Loss Scenario

Tipped (p)	Not Tipped (not-p)	Won (q)	Did Not Win (not-q)	No Tip Given
65	15	73	25	9
81%	19%	91%	31%	11%

N = 16

Table 5.5 Results for The Frequentist Win Scenario

The results for Experiment 2b can be seen below:

Percentage Frequencies of Card Selections in the Frequentist Win Scenario

Tipped (p)	Not Tipped (not-p)	Won (q)	Did Not Win (not-q)	No Tip Given
87	17	85	54	10
83%	16%	81%	51%	10%

N = 21

Combinations of selections were not produced due to the nature of the large array selection task making such a computation impossible. However, indices were produced, again based on the procedure used by Pollard and Evans, although modified to account for the task. The formula used here was (p + q + not-q) – (q + no tip). This formula produces a range of –10 to 15, again this leads to some clouding of the data, but has the advantage of giving cross condition comparisons. A graph showing these scores for both groups is shown below (Chart 5.3). A one way analysis of variance found that there was no significant difference between the two groups on this index ($F(1,35) = .283$, $p = ns$). Again this index will be referred to as the rational index.

5.8 Discussion

It can be seen from tables 5.4 and 5.5 above, and from the results of the statistical test used that there was no difference between the two scenarios. The prediction was that participants in the loss scenario would select a higher frequency of not-q cards than those participants in the gain scenario, because the former stood to lose rather than gain, and previous research would suggest that participants are more away of potential losses than potential gains. However, these predictions were not upheld as no differences were found.

Previous research would suggest that participants should show improved performance in the loss condition, as they should want to avoid losing. However, this was not the case. It is possible that the scenario did not stress the losses enough or make it clear enough to the participants what was at stake. The results showed slightly better performance in the scenario that stressed the importance of winning, although this difference was not significant. What is perhaps of more direct relevance overall however, is the influence of a

probabilistic manipulation of the task, as this is the central theme of this thesis - are participants more likely to select the not-q card when trying to assess the likelihood of q given p?

With regard to the predictions vis-à-vis the cards which the participants would choose, it can be seen that there is an increase in selection of the not-q card when compared to a standard abstract indicative task, as discussed in Chapter Two. It is argued here that there is a role for probability in reasoning, as many participants are able to see the importance of selecting the not-q card in order to make a sound probability judgment. Hence the increase that is shown in the selection of the not-q card. However, the selection of the not-q card is not up to the level of the selection of the p and q cards (p and q cards are chosen at around the 80% level, whereas the not-q card selections were 31% and 51% for the loss and gain scenarios respectively). The not-q card should be regarded as being as important in making a probability judgement as the p and q cards, and hence should show a similar selection rate. Although the rate of selection of the not-q card is not at the level of the p and q cards - it is higher than would normally be predicted for an algebraic version of the selection task. The version here uses thematic content, although the facilitation of the not-q card selection requires more than this as was demonstrated by Manktelow and Evans (1979). Thus it is argued here that requiring the participants to assess the likelihood of q given p increases the selection of the not-q card.

To follow on from this research there was a revised version of the Win scenario presented to participants, which stressed the importance of the participants winning the bet. This scenario was selected because it produced the highest level of not-q selections.

5.9 Revised Large Array Selection Task

As was discussed above the experiments so far do suggest some support for the idea that probability plays a role in reasoning with indicative conditionals. This was seen to be the case due to participants' selections of the not-q card rising up to 51% (experiment 2b Table 5.5), compared to far lower levels on the standard abstract task, where selection rate is typically around 10% (see Evans, Newstead and Byrne, 1993, for more details).

The scenario from experiment 2b that stressed to participants the importance of winning a bet was the scenario that elicited the highest selection of the not-q card, and so another gain scenario was designed. The cards with which the participants were presented were revised slightly, to make the task as clear as possible to the participants, and also the card which had previously read 'Did Not Win' was changed to read 'Lost'. Although 'Did Not Win' is an explicit negative of 'Won', and this tends to yield increased performance it was felt that 'Lost' may elicit a greater response, as it reinforces not having won, by emphasising that 'Did Not Win' is more than just failing to gain.

5.10 Experiment 3.

Method

5.10.1 Participants.

Eleven first year psychology students enrolled at the University of Wolverhampton, none of whom had any experience of the selection task. All participants participated on an unpaid volunteer basis.

5.10.2 Materials.

The scenario that was used in this experiment was as follows:

Highly Desirable Gain:

You have decided to undertake an MSc course in order to further your academic standing and career prospects. However you have been told that the likelihood of you being able to receive funding for living expenses or for your course fees is extremely unlikely.

You really do want to take this MSc course but feel that you cannot possibly do this without the financial backing you will need. You therefore decide to find some way of making money quickly so that you can study for your masters degree. Fortunately a friend tells you about the tipster in "Racing Weekly", she states that **If Racing Weekly tips a horse then it wins.** You decide that horse racing is as good a way as any to make money quickly so you decide to follow your friend's advice. However, you want to be sure that you start winning immediately and so you wish to check the reliability of Racing Weekly's tipster.

Your friend has given you her betting slips on which she has recorded details of horses she backed in different races. Each slip says whether the horse was tipped by Racing Weekly or not on one side, and on the other side whether the horse won or not.

Which of the slips would you need to turn over in order to judge whether your friend's statement will be likely to lead you to winning your bet? Please tick the appropriate answer box(es).

5.10.3. Design.

Again it was predicted that the participants should select the p, q and not-q cards, and that selection rates for the not-q card should be similar to that of the p and q cards. There were three card orders presented to the participants, and these different orders were handed out to the participants randomly. (Response sheets can be found in Appendix 3).

5.10.4 Procedure.

The task was presented to participants during their normal classes. The participants were required not to confer, and the experimenter remained with the participants to ensure that this was the case. The participants were given as long as they required for the task, and most participants completed the task within ten minutes.

5.11 Results

Table 5.6 Results for The Revised Frequentist Win Scenario

The results for Experiment 3 can be seen below:

Tipped (p)	Not Tipped (not-p)	Won (q)	Did Not Win (not-q)
35	10	40	25
64%	18%	73%	46%

The results from this experiment can be seen in table 5.6. Paired sample t-tests revealed that there was no significant difference between q card and not-q card selections ($t = 1$, $df = 10$, $p = .341$) or p cards and not-q card selections ($t = 1.936$, $df = 10$, $p = .082$).

5.12 Discussion

It is clear from table 5.6 that the revised version of the task did not seem to make any great difference to the selections which the participants made on the task, the results were roughly comparable to the results found on the original desirable gain scenario. The most remarkable result here was the drop in selections of the p and q cards, down to 64% and 73%, compared with results in the 80% level reported in the original desirable gain scenario. Therefore, whilst the selections of the p and not-q cards were not significantly different to the not-q card, this must be treated with caution, due to the low selection levels of the p and q cards.

From these results, and those presented earlier it is argued that probability does play some role in indicative reasoning, and participants do select the not-q card more than a standard abstract task when they are required to make a probability judgment. Participants see the importance of selecting the not-q card when they have to abstract some information from the cards in order to make a judgment about the likelihood of q given p. However, it is clear that although the selection of the not-q card is as important to the judgment of the likelihood of q given p as are the p and q cards the selection of the not-q card is not up to the level of the p and q cards which the participants mainly select.

Therefore, from the three experiments outlined above it can be argued that probabilistic manipulations do have some role in indicative tasks. However the effect of probabilistic manipulations was not as strong as had been expected. Selections of the not-q cards did not come up to the same level as selections for the q and p cards. As was discussed above, so called logical performance is low on standard indicative tasks (Wason, 1966, 1968), and so

it is hardly surprising that the effect of probability was not as great as had been expected. It was argued earlier that utility plays a role in deontic reasoning as demonstrated by Manktelow and Over (1991), and that the utility is a construct traditionally associated with decision-making. Following Manktelow and Over (1991) the possibility of a role for probability was investigated here, initially with indicative conditionals. It is a natural step then to return to the 'source' of the current research by investigating deontic reasoning, and the role that probability plays here, if any. There is a strong possibility that probability would play a greater role in deontic tasks, where it would integrate with the utility and search for violations elements of deontic tasks.

Therefore, the methods that have been employed here: the Large Array Selection Task, and the probability information will now be applied to deontic tasks to see the role that probability has in mediating deontic reasoning, if any.

CHAPTER 6

Page

Page

6.1 Probability and Deontic Reasoning

6.1.1. Deontic Reasoning

Over and Manktelow (1993) have defined deontic reasoning as: reasoning about the actions that one may, should or ought to perform; it is therefore, a more practical form of reasoning than reasoning about indicative conditionals. Reasoning about indicative conditionals is concerned about what may or may not be the case, and is, by its nature a more theoretical form of reasoning. That is not to say that these two forms of reasoning are separate; quite the contrary, our theoretical reasoning may have a great deal to do with any subsequent practical reasoning that humans may engage in. Thus what we do may be influenced by what we think.

It is quite plausible that probability does have a role in deontic reasoning. It is reasonable to expect a would be reasoner to consider the possible outcomes from the actions that there are to choose from. In doing this a reasoner should consider the usefulness these outcomes have to the reasoner (utility, which has been shown to play a role in the choices that participants make), and also the likelihood of the possible outcomes occurring, that is their probability. Given that decision-making appeals to the idea of *expected* utility then it is conceivable that probability will have a role in deontic reasoning.

6.2 The Probabilistic LAST

6.2.1. Probability and the Immigration Task

A task that is well known throughout the deontic reasoning literature was taken, and altered so as to contain a probabilistic element. Participants' responses were recorded, and the role

that the probability element of the task had on the choices that the participants made could be clearly examined. The task that was used here was the Immigration task of Cheng and Holyoak (1985). The paper in which this task was first presented is quite possibly one of the most important for deontic reasoning research. Not only was the paper very important *per se,* but it also prompted a vast amount of research in the area of deontic reasoning, which has seen the field develop considerably the ten years following the publication of this paper. Over and Manktelow (1995) referred to Cheng and Holyoak's paper as 'seminal'. It also saw the introduction of a new theory of reasoning, that of pragmatic reasoning schemas, which was discussed in detail in Chapter 3.

It was also decided that the frequentist framework that was used in the experiments 2 and 3 should be kept for these experiments for the same reasons that saw its introduction in the first place. The LAST format of the task presents an ideal opportunity to introduce a probability manipulation into the immigration task. In order to investigate the role of probability, it was decided that probabilistic information would take the form of the passengers' country of origin, which would be one of two possible categories, namely, European or Tropical.

In the present task the participants were given the rule *If a person has ENTERING on one side of their immigration card then they must have CHOLERA on the reverse side.* The participants were then presented with 16 cards, which contained more information than in the original version of the task. The ENTERING and TRANSIT cards that were used in the task were changed so that they told the participants both the status of the passenger, and also the country of origin of the passenger. There were two types of passenger, those who came from a Tropical country, and those who came from a European country. It was the country

of origin that was to provide the probability manipulation in the task, in that the participants would consider that a passenger from a Tropical country would be more likely to be a carrier of Cholera. Participants were given information in the scenario that would reinforce this belief.

It was predicted that the participants would show a bias to select those cards which showed that the participants had come from a Tropical country (Brazil, Thailand) compared with those cards which showed that the participants had come from a European country (Germany, Holland, Norway).

6.3 Experiment Four

Method

6.3.1 Participants.

Twenty first-year psychology undergraduates from the University of Wolverhampton participated in the task, on an unpaid volunteer basis. None of the participants had any previous experience of the selection task in any form.

6.3.2 Materials.

As was mentioned above the materials used here were an adapted version of the task used by Cheng and Holyoak (1985). For this experiment the scenario was as follows:

> Please imagine that you are working as a customs official at Heathrow airport, and as part of your duties you must check passengers' immigration cards. One side of the immigration shows whether a passenger is entering Great Britain or if they are just in transit between planes, together with the traveler's country of origin; while the other side of the card shows a list of diseases against which the traveler has been vaccinated.

You are particularly concerned that people infected with cholera should not be allowed to enter the country. It is well known that cholera is particularly common in Tropical countries.

The rule that was given to the participants was *If a person has ENTERING on one side of their immigration card then they must have CHOLERA on the reverse side.*

As can be seen from the scenario above a rationale was included (shown in italics here, though not in the original task presented to participants), this was for two reasons. Firstly it was to point out the utility for detecting possible violations, and secondly, to point out some countries have a greater probability of having passengers who are infected with cholera.

6.3.3 Design.

As mentioned earlier, the Large Array Selection Task was kept in this experiment, and the participants were presented with twenty cards, which were made up in the following way: Six were ENTERING (p) cards three of which were European cards (marked as showing that the passenger was of European origin), and three of which were Tropical cards (showing that the passenger had come from a Tropical country), four were TRANSIT (not-p) cards (two European, and two Tropical) five included "cholera" (q) and five did not include "cholera" (not-q) among the list of diseases (this response sheet can be found in the appendix 4).

6.3.4 Procedure.

The task was administered to participants during their usual classes. They were asked not to confer, and the experimenter remained with the participants to ensure that this did not happen. There were four randomised card orders, and these were administered randomly to participants. Participants were given as long as they needed to complete the task, although as in previous experiments, most participants were finished within ten minutes. Finally, at the end of the experiment the participants were thanked for their participation in the task, and given a short explanation of the broad aims of the experiment.

6.4 Results and Discussion

Table 6.1 shows the results for this experiment, in terms of percentage frequencies of selections for the cards in this task.

Table 6.1: Percentage Frequencies of Card Selections in the LAST Immigration Task.

Entering Trop (p)	Entering Euro (p)	Transit (p) (pooled)	Cholera (q)	No Cholera (not-q)
98%	57%	6%	5%	67%

N = 20

The selections for the not-p card were pooled across the two types of passenger (Tropical and European) because the selections of this card were so low, there is no benefit in listing them separately.

Participants were classified in terms of the numbers of Tropical and European p cards they selected. Nine participants selected more Tropical p cards, 11 chose equal numbers of both types of p card and none chose more European than Tropical p cards ($p = 0.002$, one-tailed sign test).

It is possible to conclude from this that the prediction that the participants would choose a greater number of Tropical p cards than European p cards was upheld. The only difference between the two types of card was the information about the origin of the passenger.

However, it is not easy to compare these results directly with those of Cheng and Holyoak, as they do not give a precise frequency of card selection in their task, and also they used a

standard four-card task, rather than the large array that was used here. However, reading from the graph that the authors gave, it appears that around 90% of their participants chose the p and not-q cards. It appears that their participants gave a higher 'correct' response than those in this experiment. While the p scores were in line with Cheng and Holyoak's (1985) results the not-q scores were lower. Hence it is possible that in the present experiment Tropical scores were raised, or European scores were suppressed. Given the design differences between this experiment and Cheng and Holyoak's, a baseline measure of LAST performance is needed. Consequently, the next experiment was a baseline measure, the Large Array Selection Task was used, but the probability information was removed from the cards. This experiment is presented below:

6.5 Baseline Measures When Using the Large Array Selection Task.

6.5.1 Rationale for Experiment 5

As was mentioned above, it was necessary to run a baseline condition in order to determine whether the results from the previous experiment were a due to there being an increase in the selection of the Tropical ENTERING cards, or a suppression of the European ENTERING cards.

In order to measure this it was necessary to run an experiment in which the probability information on the p and not-p cards was removed, that is, the information about the country of origin of the passenger. In addition to this, the effect of the rationale sentence was also tested in this experiment by including or omitting it. Cheng and Holyoak (1985) showed the importance of a rationale sentence in the facilitation of correct responding on the task, in that a higher level of correct responses occurred when the rationale sentence was present than when it was not.

6.6 Experiment Five.

Method.

6.6.1Participants.

Twenty-one first-year psychology undergraduates from the University of Wolverhampton participated in the task, on an unpaid volunteer basis. None of the participants had any previous experience of the selection task in any form.

6.6.2 Materials

As was mentioned above the rationale sentence was omitted for some of the participants, and included for others, however, the scenario given to the participants was the same for both groups, and this was as follows:

> Please imagine that you are working as a customs official at Heathrow airport, and as part of your duties you must check passengers' immigration cards. One side of the immigration shows whether a passenger is entering Great Britain or if they are just in transit between planes; while the other side of the card shows a list of diseases against which the traveler has been vaccinated.

Ten of the participants used received this information alone, and then the task, whereas eleven of the participants used in this experiment received the rationale passage as well, that was:

You are particularly concerned that people infected with cholera should not be allowed to enter the country. It is well known that cholera is particularly common in Tropical countries.

6.6.3 Design.

The task presented to the participants in this experiment was the same as that which was used in the previous experiment, except that the probability information was removed from the p and not-p cards. The cards stated whether the passenger was ENTERING the country, or just in TRANSIT between planes, as with the first of this set of experiments the q and not-q cards stated what vaccinations the passengers had been given prior to them beginning their journey.

Obviously, the Large Array format of the task was retained in this experiment as it was designed to be a comparison to the previous experiment, the overall format of the task was also retained, as the participants were presented with the same combinations of cards as they were originally. That is, six p cards (ENTERING), four not-p cards (TRANSIT), five q cards (CHOLERA absent), and five not-q cards (CHOLERA present).

Again four random card orders were used in this experiment, so as to counterbalance against any effect of card orders. (An example of the response sheet used here can be found in appendix 5)

6.6.4 Procedure.

Participants were allocated at random to one of the two groups, that is rationale present, or rationale absent, and the experiment was run during one of the participants' normal classes. The participants were asked not to confer, and the experimenter remained present to

ensure that this was the case. Participants were told that they had as long as they required to complete the task, although most were finished within ten minutes. When the participants had all completed the task they were thanked for their participation in the task, and given a brief explanation of the broad aims of the experiment.

6.7 Results and Discussion

The percentage selection frequencies of cards for the two groups can be seen in tables 6.2 and 6.3 below.

As was stated above the main aim of this experiment was to study the selection of the p cards, and to compare these to such selections in the original frequentist immigration task, presented in table 6.1. It is clear from the tables 6.2 and 6.3 below that the selections of the p cards here compare closely with the frequency of selections of the Tropical p cards in the original experiment.

Table 6.2 Percentage Frequencies of Card Selections in the LAST Immigration Baseline Task with Rationale.

Entering (p)	Transit (not-p)	Cholera (q)	No Cholera (not-q)
80%	9%	27%	53%

N = 11

Table 6.3 Percentage Frequencies of Card Selections in the LAST Immigration Baseline Task without Rationale.

Entering (p)	Transit (not-p)	Cholera (q)	No Cholera (not-q)
98%	2%	45%	50%

N = 10

Hence it seems to be that the probability manipulation in the original experiment had the effect of suppressing the selections of the low probability (European ENTERING) cards, rather than facilitating selections of the high probability (Tropical ENTERING) cards. Thus, it is suggested here that the participants saw those passengers who are coming from European countries (low-risk areas) as being less subjectively significant. This claim can also be corroborated by the findings that the p card has a normally high selection rate, ignoring cases which manipulate utility (such as Manktelow and Over, 1991 and Cosmides, 1989).

Finally, to consider briefly the effect of the rationale sentence of the selections which the participants made. In Cheng and Holyoak (1985), the rationale sentence led to a facilitation of p and not-q selections. However, this did not appear to be the case here, as the not-q frequency is 53% and 50% for the rationale and no rationale conditions respectively. Selections of the p card were lower in the rationale condition than in the no rationale condition, although this was only a slight difference. It is also worthy of note that the frequency of q card selections was quite high here, and certainly higher than in the original experiment, this is put down to the lack of a clear connection between a rationale and specific probability and utility information, thus producing more variable performance. It is also important to consider what effect the Large Array Selection Task may have on the choices which the participants make, given that they are presented with six times as many cards as in the standard four card version of the task. There needs to be more research to study the connection between selections of the four-card task, and the Large Array version of the task, not only in the deontic domain here, but also with indicative conditionals, and abstract materials. However, participants here did seem to produce the predicted pattern like Cheng and Holyoak (1985).

In order to further examine the effect of the rationale sentence of the selections which the participants made a version of the task was run in which the rationale was again omitted but the information about the country of origin on the p and not-p cards was reinstated. This experiment is presented below.

6.8 The Effect of Rationale Sentences on Large Array Deontic Conditionals

6.8.1 Motivation for Experiment Six

As was discussed earlier the effect of the rationale sentence needed to be examined in more detail. Cheng and Holyoak (1985) showed that when participants were provided with a rationale for a rule, then they could solve the task reasonably well, using Johnson-Laird, Legrenzi and Legrenzi's (1972) postal task. This was the aim of the experiment presented here. The rationale sentence that was used for some of the participants in the experiment five was omitted for all the participants in this experiment, but the probability information on the p and not-p cards was reinstated, that is the information about the country of origin of the passengers. Thus the main aim of this experiment was to examine whether there needs to be a link between an explicit rationale setting out the probability and utility structure of the scenario and corresponding information in the problem content.

It is already known that the rationale passage that has been used in the previous experiments contains both utility and probability elements. That is that the first part of the rationale states that the authorities are particularly concerned to keep people infected with

cholera out of their country, and the second part states that this disease is particularly *common* in Tropical countries.

This experiment was designed to test the hypothesis that the participants need to be explicitly cued to this rationale, and that they are not able to construct it for themselves. Thus the rationale sentence was completely omitted, and the country of origin information was replaced on the p and not-p cards, as in the original experiment that was presented above.

6.9 Experiment Six.

6.9.1 Participants. Fifteen first year psychology undergraduates from the University of Wolverhampton participated in the task, on an unpaid volunteer basis. None of the participants had any prior experience of the selection task in any form.

6.9.2 Materials

The participants were presented with the following scenario:

> Please imagine that you are working as a customs official at Heathrow airport, and as part of your duties you must check passengers' immigration cards. One side of the immigration shows whether a passenger is entering Great Britain or if they are just in transit between planes, together with the traveler's country of origin; while the other side of the card shows a list of diseases against which the traveler has been vaccinated.

6.9.3 Design

However, there was no rationale sentence given to the participants at all, and this was the only instruction that the participants were given. Obviously participants were told to place a tick in the box corresponding to those cards which the participants thought should be turned over to ensure that the rule was not being violated. The rule that the participants were given was the same as the original rule given to the participants in experiment 4, that is *'If a person has ENTERING on one side of their immigration card then they must have CHOLERA on the reverse side'*. The participants were then presented with a piece of A4 paper with the twenty cards on that had been presented to the participants in the original experiment, containing the probabilistic information on the p and not-p cards, unlike those cards presented to the participants in experiment five. There were four randomised card orders used to ensure that there was no effect of card orders on the selections that the participants made. (The response sheet here was the same as that for experiment four, and can be found in Appendix 4).

6.9.4 Procedure

The task was administered to the participants during their regular classes, and participants were asked not to confer, the experimenter remaining present to ensure that this was the case. The four randomised card orders were administered randomly to the participants. The participants were allowed as long as they needed to complete the task, although most of the participants had completed the task within ten minutes. When all participants had completed the task the participants were given a brief explanation of the broad aims of the experiment, and thanked for their participation in the experiment.

6.10 Results and Discussion

The percentage selection frequencies of cards for experiment 6 can be seen in table 6.4 below.

Table 6.4 Percentage Frequency of Selection of Cards in LAST Immigration Task, with no Rationale

Entering Trop (p)	Entering Euro (p)	Transit (p) (pooled)	Cholera (q)	No Cholera (not-q)
93%	93%	2%	48%	33%

N = 15

Table 6.4 shows the percentage selection frequencies of all cards for all participants, again the frequency for the not-p cards has been pooled across European and Tropical cards, because the selection was so low for these cards.

Regarding selection of the European and Tropical p cards the results are as clear as they possibly could be, that is that the frequencies are exactly the same. The frequency of selection of these cards is at the level of the Tropical p cards from the first experiment, or the "non-suppressed" level. The return of selections to the non-suppressed level clearly indicates that the effect of the probability information on the cards in experiment four was to suppress selections of European p cards rather than to facilitate selection of Tropical p cards.

The trend shown in experiment five was continued here, in that the difference between the q and not-q cards was not as clear as it could be. It therefore appears that there must be explicit representations of utility for at least one of the parties involved in a deontic reasoning task are required for clear patterns of inference to be displayed (Manktelow and Over, 1992). It seems from the results from this experiment that the same goes for probability, in that it needs to be explicitly represented for the expected patterns of inference to be observed.

There is one final experiment to be reported in this chapter. An experiment in which the participants were given the probability information, the country of origin of the passenger, but this time it was on the q and not-q cards, rather than on the p and not-p cards. Therefore, the final experiment in this set places this information on the consequent cards, and investigates whether this will have an affect on the choices the participants will make.

6.11 The Effect of Probability Information on Consequent Cards

6.11.1 Rationale for Experiment Seven

As was stated in 6.9 the experiment here is the same as experiment four, except here the probability information was presented on the q and not-q cards rather than on the p and not-p cards. There is some work to suggest that probability judgments can affect such inferences (Byrne, 1989 and Stevenson and Over, 1995). Byrne (1989) found that additional premises had an effect on the corresponding modus tollens inference in indicative conditional syllogisms. However, it is not possible to directly link this to deontic reasoning, as it is not possible to refer to the modus tollens and modus ponens inferences in an indicative task as being equivalent to the corresponding inferences in a deontic task. Deontic conditionals are not truth-functional, unlike indicative conditionals, this is clearly indicated by the use of modal auxiliaries in their consequents.

Therefore, although there is not necessarily a link between inferences that participants make on indicative and deontic conditionals, predictions can still be made. It is predicted that even when the information about the passenger's origin is on the q and not-q cards, which results in the participants being able to see both the country of origin, and the vaccinations which the passengers have had on the same card, they will still show a bias. It is hypothesised that the bias will be a suppression of those cards that show that the passenger has come from a European country.

6.12 Experiment Seven.

Method

6.12.1 Participants

Twenty-one first and second year psychology undergraduates from the University of Wolverhampton participated in the experiment, on an unpaid volunteer basis. None of the participants had any experience of the selection task in any form.

6.12.2 Materials.

Obviously it was necessary to make some fairly major changes to the instructions that the participants were given in this experiment. The scenario with which the participants were presented was as follows:

> Please imagine that you are working as a customs official at Heathrow airport, and as part of your duties you must check passengers' immigration cards. One side of the immigration card shows whether a passenger is entering Great Britain or if they are just in transit between planes; while the other side of the card shows a list of diseases against which the traveler has been vaccinated, together with the traveler's country of origin.
>
> You are particularly concerned that people infected with cholera should not be allowed to enter the country. It is well known that cholera is particularly common in Tropical countries.

Again the participants were given the rule, that *If a person has ENTERING on one side of their immigration card then they must have CHOLERA on the reverse side.*

6.12.3 Design

The participants were then presented with twenty-two cards, this was because there were other cards added to the response sheet in this experiment. Two cards were added, these were one q card, and one not q card, making six in all. Thus it was possible to make three of the q cards European, and three as Tropical, and the same with the not-q cards. The other cards were set at the same level as the original experiment, that is four not-p cards, and six p cards. Obviously, the probability information was removed from the p and not-p cards for this experiment, and was only present on the q and not-q cards. (The response sheet can be found in the appendix 6).

6.12.4 Procedure

The task was administered to participants during their usual classes, and they were asked not to confer, the experimenter remained with the participants to ensure that this did not happen. There were four randomised card orders, and these were administered randomly to participants. Participants were given as long as they needed to complete the task, although as in previous experiments, most participants were finished within ten minutes. Finally, at the end of the experiment the participants were thanked for their participation in the task, and given a short explanation of the broad aims of the experiment.

6.13 Results and Discussion

The percentage selection frequencies of cards for experiment 7 can be seen in Table 6.5 below.

Table 6.5 Percentage Frequency of Selection of Cards in LAST Immigration Task, with no Rationale

Entering (p)	Transit (not-p)	Cholera Trop (q)	Cholera Euro (q)	No Cholera Trop (not-q)	No Cholera Euro (not-q)
96%	1%	16%	13%	67%	40%

N = 21

Again, the results from this experiment are clear, and in the direction that was predicted earlier: Participants show a suppression of those cards which show that the passenger has come from a European country.

The results from experiment 7 are quite remarkable considering that the participants can see not only what country the traveler is coming from but also what vaccinations the traveler has had, on the same side of the card. When the participants are classified as to the types of card that they chose the following results were found: six chose more Tropical not-q cards than European not-q cards, none chose more European not-q cards than Tropical not-q cards, and the remainder of the participants chose equal numbers of not-q cards (P=0.016, one-tailed sign test).

When comparing these not-q card results to those in the original immigration experiment, it can be seen that again the effect of the probability manipulation is to suppress selection

of the not-q cards for European travelers, rather than facilitating selection of the not-q card for the Tropical travelers.

Finally, the level of q card selection has dropped back to the original level, rather than those levels which were found in the interim experiments (around 50%). This supports the suggestion that was made earlier that there needs to be a clear link between task information and task materials for the participants to ignore those cards that are of little importance. As experiments 5 and 6 omitted part of the instructions or task materials the participants did not eliminate the q cards from examination, which, of course, is fruitless as it is unable to reveal any violations.

6.14 Chapter Summary

Before, moving on to the final set of experiments in this thesis, there is scope for a general overview of the findings of the experiments that have looked at the role of probability in deontic reasoning. Other issues to be addressed here include how these findings relate to the first set of experiments that used indicative conditionals.

The initial set of experiments suggested that there was some evidence for participants selecting not-q when trying to assess the likelihood of q given p. The not-q card is typically ignored in standard indicative versions of the task, and yet selection rates here rose to approximately 50% in some cases. Manktelow and Over (1991) demonstrated a role for utility in deontic reasoning, a construct was usually associated with decision-making. It was a natural step then to consider a further construct from this area: probability. Probability was initially considered alone using indicative conditionals (the horse race tasks, experiments 1-3), and some effect was found, as discussed above.

One of the major outcomes of this shift to consider probability was the introduction of the Large Array Selection Task (LAST). This afforded a number of benefits over the standard four-card task. It allows participants to reflect the relevant importance of each card in the task. For example, when presenting with 5 not-q cards a participant can select 0, 1, 2, 3, 4, or all of these depending on how relevant they believe that card to be to the task. This flexibility is not present in the standard task where a card is simply selected or not. The task also allows a large amount of data to be collected using reasonably few participants. Finally, it helps to avoid the cognitive illusions cited by Gigerenzer (1991) by presenting the task in a frequentist framework.

The horse race task provided some basic evidence for the use of probabilistic information by participants when presented with the task. The basis for studying the role of probability had been Manktelow and Over's (1991) findings with utility, hence it was a reasonable step to return to deontic reasoning for the second set of experiments, using Cheng and Holyoak's deontic immigration task.

Deontic tasks facilitate selection of the 'correct' cards (dependent on the utility and social role of the participant), it seemed to be a prudent move to design a task which involved the study of probability through the medium of a deontic task. Thus a version of the immigration task originally presented by Cheng and Holyoak (1985) was used, in which the origin of the passengers was also put on to the cards. This had a marked effect on the choices that the participants made, showing a bias against selecting those cards that showed that the passenger had come from a European country.

There is scope here for a comparison between the two sets of experiments conducted thus far. The indicative experiments suggested that probabilistic information might be important in reasoning with conditionals. The move to deontic reasoning confirmed this - with the suppression effect being very clear when participants do not consider a passenger to be a threat to the rule they have been given. Thus the probability information appeared to have a greater effect when the rules the participants were given were deontic rather than indicative, as was predicted earlier.

Subsequent experiments showed that the effect of the probability manipulation was to suppress the selection of European cards rather than to facilitate the selection of the Tropical cards. Suggesting that the participants judge the Tropical p cards as a greater risk of being a violation of the rule that the participants were given rather than the European cards. The effect would need to be one of suppression rather than facilitation for a number of reasons. Firstly participants on the task almost universally select the p card, and thus the only shift could be a decrease in selections or a suppression effect. Secondly, all passengers represent a risk, however, some are less risky than others (those from European countries, where Cholera is not particularly prevalent). Thus this should be reflected in a suppression effect. If all passengers have a 'default classification' of risky, as would be suggested by Cheng and Holyoak's (1985) results as well as the experiments reported here with the rationale sentence removed, then the only shift that can occur to this 'classification' is to reduce the risk rating of any type of passenger.

It also seemed to be the case that participants require a clear relation between the task instructions, and the task materials, and that the participants need to have the probabilistic aspects of the task explicitly pointed out to them. When this information was not made explicit the participants selected both types of p card at exactly the same level. Finally, the participants displayed the suppression effect even when the probability information was put onto the consequent cards and the participants could assess both the country of origin, and the vaccinations that the passengers' had had. There will be further discussion of the results of these experiments in the general discussion, where theories of both indicative and deontic reasoning will be considered in light of the findings presented here.

The final set of experiments was designed to examine *how explicit* information about probability needs to be for participants to react to it. Manktelow and Over (1991) suggested that information about utility and social roles needed to be explicit for participants to select accordingly on the task. Thus experiments were designed which would require the participants to extract information about probability. Their performance on a probabilistic selection task was then assessed. Participants would be presented with information about passengers that showed one type (European or Tropical) of passenger violated the rule whereas another did not. They were then given the LAST where they should display knowledge of this bias if they have extracted the probabilistic information from the initial cards with which they were presented.

CHAPTER 7

Page

7.1 The Computer Based Selection Task

7.1.1 Rationale for Experiment 8

This experiment was a computer based large array selection task, which allowed the participants to click on the cards, and examine the reverse side of the card, which usually remains hidden. In this version of the task the participants receive more information about the nature of the cards. They can see whether they conform to the rule or not, thus they have more information than is usually given to participants in regular pencil and paper tasks - such as those used in other experiments presented both in this thesis, and the majority of the selection task literature.

The proposal behind this experiment was to present different information on the cards to three different groups, and to see how this information would influence the choices that the participants subsequently made as they progressed through the task. The participants begin their task, seeing one type of card as more likely to be a violation than another. From the previous experiments it can be seen that participants make the choice that Tropical ENTERING cards constitute a greater risk to the violation of the country's law on immigration than European ENTERING cards do. Subsequently they display a suppression effect on the European p cards when compared to the Tropical p cards. Thus the participants must have a certain model of the task in their minds before they begin. However, if they subsequently receive information that confounds these models then they should search for an alternative model that fits with this new data.

Thus the problem content was kept broadly the same, in that the Large Array Immigration Selection Task was still used, and three groups were used, group number one: in which no bias was shown on the cards; there was an equal number of violation of both European and Tropical cards. Group two in which there was a bias of European violations; all the European ENTERING cards were violations of the rule; they showed ENTERING on the one side of their form, but they did not show CHOLERA on the other. Finally, Group three: in which there was a bias of Tropical violations; all the Tropical ENTERING cards were violations of the rule, for the same reason as given above for the European violations.

It was predicted that the participants would extract the information from the immigration cards as they passed through the task, and that they would show a bias towards the cards that showed a bias in their presentation. More explicitly, group number one would show a suppression of selection of the European ENTERING cards, this is because even though there is no bias in the cards to begin with, the participants should still show the suppression effect that was observed in experiment 4. Group two, however, should extract the information from the cards with which they are presented that the most likely passengers to violate the rule are those travelling from a European country. This should facilitate selection of these cards bringing them up to a similar level to the Tropical ENTERING cards. The third group should show an even greater suppression effect than that which has been previously observed, as their original models were being reinforced by the information that they receive on the cards with which they were presented.

7.2 Experiment Eight

Method

7.2.1 Participants

Forty-eight first year psychology undergraduates from the University of Wolverhampton participated in the task, on an unpaid volunteer basis. None of the participants had any previous experience of the selection task in any form.

7.2.2 Design and Materials. In this task the participants were initially presented with an instruction screen, and then with two pages of immigration cards, each containing 16 cards (thus 32 in all). Again four randomised card orders were used, to eliminate any possible effect of card orders. ("Screen Dumps" of the pages with which the subject were presented can be found in appendix 7).

The task was generated using the Macintosh HyperCard software package, with the participants having to move the mouse to the card which they thought needed to be turned over to check that the rule was not being violated. The selections that the participants made was recorded by the computer, and examined later.

The instructions that the participants were given were as follows:

Please imagine that you are working as a customs official at Heathrow airport, and as part of your duties you must check passengers immigration forms. One side of the form states whether the passenger is entering the country or just in Transit, together

with the passenger's country of origin. While the other side of the form shows a list of diseases against which the passenger has been vaccinated. You are particularly concerned that people infected with cholera should not be allowed to enter the country. It is well known that cholera is particularly prevalent in Tropical countries. You have, therefore, been given the following rule, **If a form says ENTERING on one side then CHOLERA must be among the list of diseases on the reverse side.**

On the next two pages are 32 such immigration forms, you must click the mouse on the forms that you think you must check in order to ensure that this rule is not violated, you will then be shown the reverse side of the card. You are free to return to this instruction page at any time during the experiment.

The participants were then required to 'click' on a box marked 'click here to begin the experiment', and they were then presented with the first of the two screens of cards. The participants moved on to the second screen by 'clicking' on another box, marked 'next page', and were then presented with the second screen of cards. When the participants had completed the experiment they were required to 'click' a box marked 'finished'; they were then presented with a 'thank-you' page, and told that they could now leave. Finally, at any time during the experiment the participants could return to the instructions page by 'clicking' a box marked 'return to instructions'.

7.2.3 Procedure. The participants were asked to participate in the experiments during their usual classes, and if they agreed to do so the experimenter met them after their class, and escorted them to the laboratory. The participants were asked if they had basic mouse skills,

and if they did they then began the experiment, those that did not have such skills were shown how to do use a mouse.

The participants were then shown into individual cubicles to begin the experiment, they were left in the room alone, and told that if they did not understand the instructions then they should ask the experimenter. Any such requests were met with a reiteration of the instructions given in the task, which proved to be sufficient explanation for the participants.

The participants were given as long as they needed to complete this experiment, although most participants were finished within ten minutes. When the participants had finished the task they were thanked for their participation, and given a brief outline of the broad aims of the experiment.

7.3 Results

The results for this experiment can be seen in the table 7.1 below. Predictions were upheld for the control group, although not for the two experimental groups the suppression effect was still present in the EATN (European All biased Tropical None biased) group, when it was predicted that it would diminish. However, selection rates in this group were comparable with both the control group and the ENTA (European None biased Tropical All biased) group.

Percentages of Card Selections For The Three Groups in Experiment Eight

Table 7.1

Control Group.

Entering Euro (p)	Entering Trop (p)	Transit Euro (not-p)	Transit Trop (not-p)	Cholera (q)	No Cholera (not-q)
54%	80%	24%	44%	10%	54%

N = 17

Experimental Groups.

Euro All Trop None (EATN)

Entering Euro (p)	Entering Trop (p)	Transit Euro (not-p)	Transit Trop (not-p)	Cholera (q)	No Cholera (not-q)
58%	85%	21%	58%	7%	29%

N = 13

Euro None Trop All (ENTA)

Entering Euro (p)	Entering Trop (p)	Transit Euro (not-p)	Transit Trop (not-p)	Cholera (q)	No Cholera (not-q)
64%	90%	6%	26%	30%	33%

N = 18

In order to carry out an analysis of variance on the results that were observed it was necessary to compute a falsification index for the cards that were used in the task. Using the following formula does this:

$$(p + \text{not-}q) - (\text{not-}p + q).$$

Based on Pollard's indices, Pollard and Evans (1987). Descriptive statistics for these falsification indices are given below in Table 7.2

Descriptive Statistics For Experiment 8

Table 7.2

Group	Number	Minimum	Maximum	Mean	Std. Deviation
CONTROL	17	-2	10	3.29	3.7
ENTA	18	-1	16	5.22	4.72
EATN	13	-1	14	4.62	4.09

An Arc Sine transformation was applied to the percentage card selections for each subject, so that the falsification index could then be computed, and used as the data in an analysis of variance (Snodgrass, 1977). The results from the Anova are reported in table 7.3 below:

Table 7.3

	Sum of Squares	df	Mean Square	F	Significance
Between Groups	53.128	2	26.564	3.881	.028
Within Groups	308.035	45	6.845		
Total	361.163	47			

The results from experiment eight do not support the predictions that were made. There was a significant difference between the FI values for the groups (F= 3.881, p=.028). A post-hoc Student - Newman - Keuls test was applied to the ANOVA results to see where the differences exist. This revealed that group 1 (Control) was significantly different from group 3 (ENTA).

This translates as there being a greater selection of logically correct cards by the participants in group 3 than in group 1. This could mean that the participants in the ENTA group were tending to select the p and not-q cards more than those participants in the control group. However, a lower not-p card selection would also raise the index, and it appears that this was the case. There was no difference between the ENTA group and the EATN group as had been originally predicted.

7.4 Discussion

The first thing to note is that generally speaking the results are not as clear cut as had been predicted above. The groups did not show the biases in their selections that should have been suggested to them from the information that they were given on the cards in the earlier part of the experiment.

If the ENTERING cards are examined only then it can be seen that there is no large difference in the percentage frequencies of selections of the Tropical cards, 80%, 85%, and 90% for control, EATN (Euro All Trop None) and ENTA (Euro None Trop All) respectively. This was not what had been predicted, in that there should have been a drop in the selections of these cards in the EATN group, but this was clearly not the case. For the European ENTERING cards it is again the case that the results do not follow the predictions that were made. Here the percentage frequencies of selections were 54%, 58% and 64% for the control EATN and ENTA groups respectively. These do not follow the predictions in that the EATN group should have shown the highest selection of these cards, but this was not the case. It can be seen that there are really no huge differences between the three groups with regard to the selection of the ENTERING cards, regardless of whether they show that the passenger was from a Tropical country or from a European country. Perhaps then, it can be argued that due to the very robust nature of the original effect (see experiment four), people's prior beliefs about diseases and countries of origin are immune to any new information that they may subsequently receive. This explanation is upheld by the only significant difference being that the group which were presented with the cards which showed that the Tropical passengers were more likely to be violations (the

ENTA group) did not show any suppression of the logically correct cards, when compared to the control group.

The other cards of major concern are the No Cholera cards, the selections here do vary across the groups which is strange. Participants should really select all the cards which show 'No Cholera' as they contain no information about the origin of the passengers, but do go some way to forming a potential violation. However, the groups do show a big difference in the selections of these cards, 54%, 29% and 33% for the control, EATN and ENTA groups respectively. It is also interesting to note that selections of the q card rose to 30% in the ENTA condition, which is almost as high as this group's selection of the not-q cards, which was not as predicted.

Finally, it is interesting to note that the participants did make some unexpected selections of the TRANSIT cards. For the European TRANSIT cards the selections were 24%, 21% and 6% for the control, EATN and ENTA groups respectively, and for the Tropical TRANSIT cards the selections for the same groups were 44%, 58% and 26%. These results are again strange as there should be a low level of selection for the TRANSIT cards, but the selections of Tropical not-p cards in the EATN group were as high as the European ENTERING cards for the same group. One possible explanation is that the participants did not fully understand the nature of being in Transit, and thus they selected the cards. This may be the case as there is a bias towards selecting those TRANSIT cards which showed that the passengers were of Tropical origin, or a suppression of those which stated that the passenger was of European origin. This seems to be a fairly plausible explanation, as the Transit cards reproduce the effects observed on the ENTERING cards. However, in other versions of this task there does not appear to have been such a misunderstanding (see

earlier versions of the task in this thesis and Cheng and Holyoak, 1985), and so perhaps there is an alternative explanation for these results. It may be that because the task is computer based, and not a pencil and paper task, the participants do not feel as committed to the choices that they make on the task, as they feel that it has an air of hypothesis testing about it. Consequently, they make some choices that are non-essential just to get 'a feel' for the task. Also using the LAST, like the RAST (reduced array selection task) participants are really performing multiple selection tasks, and they may feel that they are able to examine more of the cards on such a version of the task than they do on the single shot four card task. There is an element to the LAST that allows the subject to 'revoke' an initial selection in a later presentation.

The hypothesis testing idea fits well with the work of Oaksford and Chater (1994), and their rational analysis of the selection task. However, the theory that they propose would be unable to account for the participants selecting the not-p card as much as they do, as the theory is based on information gain, and the participants do not stand to gain much information, if any, by selecting the not-p card. However, it is possible that with only 32 cards being presented to the participants there is not enough time for them to examine the information on the cards and to alter the models which they are using to solve the task, and for this change in models to show in the results from the experiments.

7.5 Revised Computer Based Selection Task

A second version of the task was designed, in which the participants were given a trial run through 48 cards - where the bias was contained. Thus, one group was presented with a set of 48 cards that contained cards where the split of violations was equal for European or

Tropical passengers (control). A second group was presented with 48 cards which had all the European passengers as violators, and none of the Tropical passengers as violators (EATN). The final group was presented with 48 cards which showed all the Tropical passengers, and none of the European passengers to be violators of the rule (ENTA). The participants were then presented with the cards 'for real', and here no bias was given: the cards had an equal split of European and Tropical violations. Thus it was possible for the participants to extract the information about the cards from the trial run, and then adjust their models, and it was expected that the difference in the groups that was predicted above would show up in the latter section of this experiment. As was mentioned in the discussion of the results above, the results were not as predicted and some speculative explanations for this were offered. In order to account for the unpredicted results that were observed, a revised version of the task was presented to participants to help to provide a better test of the original predictions.

The predictions for these different groups were that the first group (control) would show a suppression of selection of the European ENTERING cards, as had been the case in the earlier experiments. The second group was expected to show a bias towards selecting the European ENTERING cards. The cards that they were presented with had all the European ENTERING cards as violations of the rule that the participants had been given. Thus it was predicted that the participants would alter the models that they were using to 'solve' the task from selecting the Tropical ENTERING cards, to a model which would lead them to select the European ENTERING cards.

The third group were presented with a set of cards which had the Tropical ENTERING cards as violations, and it was predicted that this group would show a very strong bias

towards selecting the cards that showed that the passenger was ENTERING from a Tropical country.

Therefore the task was revised to remove some of the obstacles which had previously prevented a full test of the predictions that were outlined above. The task was extended to include a trial run of cards before the experimental cards were presented. It was felt that the participants in the previous experiments did not have enough cards to change the models that they were using and for this change to show in the results of the experiments.

Thus the participants were presented with 48 cards which showed the biases that were outlined above, the participants were told that they could 'click' on these cards and the card would turn over and the other side of the card would be revealed to the participants. All the participants were then presented with the same set of 32 cards on which there was no bias, that is, the cards showed an equal number of violations and non-violations for both the Tropical and European ENTERING cards. It was predicted that the expected biases would then appear in the second set of 32 cards.

7.6 Experiment Nine.

Method

7.6.1 Participants

Forty-two participants participated in the experiment, the participants were first year undergraduates enrolled on the psychology B.Sc. course at the University of

Wolverhampton. All were unpaid. None of the participants had any prior experience of the selection task.

7.6.2 Design and materials

The participants were initially presented with a set of instructions about the trial run which are shown below:

Please imagine that you are an immigration officer at Heathrow airport. Oneof your duties is to check passengers' form H. On one side of the form is the status of the passengers, whether they are entering the country or in transit between planes. On the other side of the form is a list of vaccinations that the passengers have had. You are told by the authorities that **If the form says Entering on one side, then it must have cholera among the list of vaccinations on the other side.** You are also told that cholera is particularly prevalent in Tropical countries. On the next page is a selection of sample forms for you to look at, when you click on the form it will turn over so that you can see the other side. When you have finished looking at these there are some more forms that you must examine, further instructions will be given when you come to this section. Thank you, you may now begin.

The participants were then presented with the 48 cards that contained the biased information regarding the passengers' countries of origin. The participants had as long to look through this information as they required. When they had completed this section the

participants were then presented with another set of instructions regarding the second set of 32 cards, these instructions were as follows:

> Again, please imagine that you are working as an immigration officer at Heathrow airport. You must now check passengers' form H for real, having looked through some sample forms, as you know, one side of the form is the status of the passenger, whether they are Entering the country, or in transit between planes. On the other side of the form is a list of vaccinations that the passengers have had. You are told by the authorities that **If the form says Entering on one side, then Cholera must be among the list of vaccinations on the reverse side.** You are also told that cholera is particularly prevalent in Tropical countries.
>
> On the next two pages are 32 such immigration forms, which you must now check for real. In order to ensure that this rule is not violated, you must click on those forms that you think you need to check in order to make sure that the above rule is not violated.
>
> You may return to this instruction page at any time, you may now begin the experiment.

The participants were then presented with 32 cards which contained no bias whatsoever, that is, there were equal numbers of European and Tropical passengers entering the country who were not vaccinated against cholera, and were thus breaking the rule that the participants had been given. The computer recorded the choices that the participants made on this second set of cards. Again the application used to design this experiment was

HyperCard on the Apple Macintosh. It is important to note that there was a slight error in the design of the card layout here, which was not discovered until the experiment had already been run. The error lay in the fact that there were 2 European TRANSIT cards, and 4 Tropical Transit cards (where there should have been 3 of each type), and there were 7 Cholera cards, and 9 No Cholera cards (where there should have been 8 of each type). Due to the fact that the proportions of cards selected will be Arc Sine corrected (Snodgrass, 1977) this design fault can be overcome statistically, by then producing a falsification index based on the Arc Sine corrected scores. However, it was felt that this design error is unlikely to have had any great effect on the results of the experiment. There are still enough Cholera and No Cholera cards for the participants to select, and the selection rates for the European TRANSIT cards has previously been relatively low. Although it must be noted here that the low number of European TRANSIT cards for the participants to select could possibly lead to a lower selection of these cards than there might otherwise have been. However, in defence of the design, the participants should already have formed the models that they are going to use by the time they arrived at this section of the task. As the previous experiment showed, subsequent information by that stage is unlikely to lead to any great changes in the selections the participants made. (Examples of the initial set of cards and the response screens can be found in appendix 8)

7.6.3 Procedure

The participants were asked if they would participate in the experiment during their seminar sessions, if they agreed to take part in the experiment then the experimenter took them to the laboratory. The participants were all tested individually, in a cubicle. The participants were asked if they had mouse skills, and if they did they were told to read the instructions, and to begin. If the participants had no mouse skills then they were shown

how to use the mouse, and then told to begin. All participants were told that if they became stuck at any point then they should ask the experimenter for clarification. Any clarification given involved a reiteration of the instructions that the participants had already been given. The participants were given as long as they required to finish the task, but as before most finished inside ten minutes. They were then thanked for their participation, and the broad aims of the experiment were explained.

7.7 Results

The results for this experiment can be seen in the table 7.4 below. The table shows that the suppression effect was not observed in the EATN group as was predicted, but was still observed in the ENTA group when compared to the control group, although it should be noted that p card selection rates were low overall for the ENTA group.

Percentages of Card Selections For The Participants In Experiment Nine

Table 7.4

Control Group.

Entering Euro (p)	Entering Trop (p)	Transit Euro (not-p)	Transit Trop (not-p)	Cholera (q)	No Cholera (not-q)
72%	88%	17%	35%	36%	42%

N = 15

Euro All Trop None.

Entering Euro (p)	Entering Trop (p)	Transit Euro (not-p)	Transit Trop (not-p)	Cholera (q)	No Cholera (not-q)
70%	77%	4%	11%	18%	71%

N=12

Euro None Trop All

Entering Euro (p)	Entering Trop (p)	Transit Euro (not-p)	Transit Trop (not-p)	Cholera (q)	No Cholera (not-q)
47%	54%	13%	18%	16%	53%

N = 15

The proportions of card selections for each card, for each subject was then Arc Sine corrected, and the formula for producing a falsification index was then applied:

$$(p+not\text{-}q)-(not\text{-}p+q).$$

An analysis of variance was then applied to the corrected FI Values for the three groups showed that there just fails to be a significant difference at the 5% significance level between the FI values of the three groups. The results from the Anova are given in Table 7.5 below:

Table 7.5

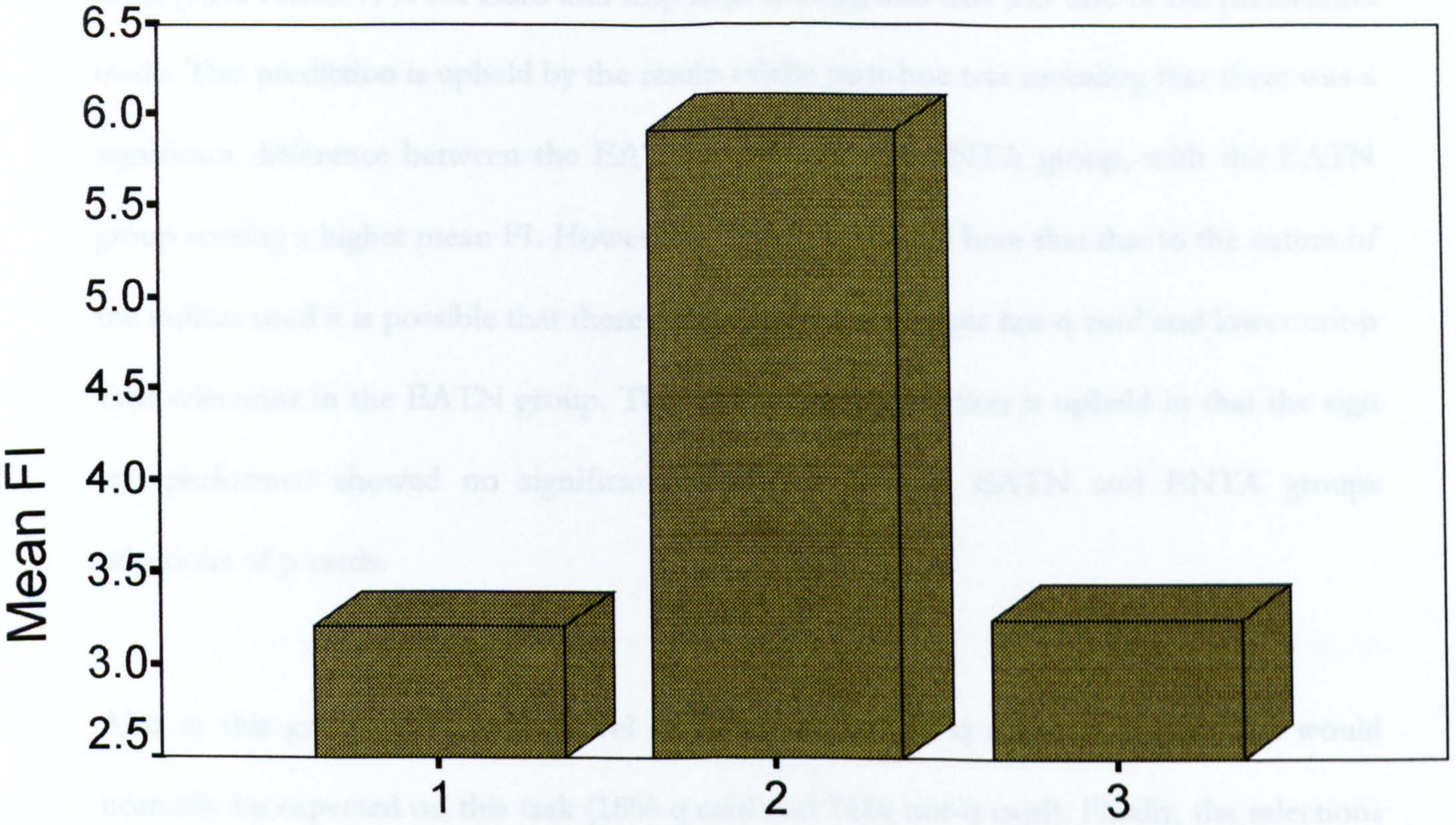
	Sum of Squares	df	Mean Square	F	Significance
Between Groups	61.795	2	30.898	3.071	.058
Within Groups	392.336	39	10.060		
Total	454.131	41			

However, application of a post-hoc Student - Newman - Keuls test with significance level 5% revealed that there was a significant difference between groups 2 and 3. These groups were EATN and ENTA respectively, with group 2 having a higher mean FI value than group 3. This does, in part support the predictions that were made, in that the suppression effect of the European p cards should reduce in the EATN condition. However, p card selections were low in the ENTA group generally, and this may account for the observed difference.

In order to examine this more closely sign tests were performed on the Tropical and Entering p cards for each of the groups to see if there was any difference here as this is where the prediction made lies. The results of these sign test were: Control: 2 participants selected more Euro than Tropical p cards, 8 selected more Tropical than Euro p cards and 5 tied (p= .109 2-tailed). EATN group: 1 participant selected more European p cards than Tropical, 3 selected more Tropical p cards than European, and 8 tied (p= .625 2-tailed). ENTA group: 2 participants selected more European than Tropical p cards, 4 selected more Tropical than European p cards and 9 tied (p= .688 two-tailed). A graph showing the mean FI values for the three groups is given in chart 7.1 below:

CHART 7.1

MEAN FI VALUES FOR EXPERIMENT 9



GROUP

1=CONTROL 2=EATN 3=ENTA

7.8 Discussion:

There is evidence from the data that something along the lines of the predictions is indeed taking place. If the percentages of selections for the EATN group are considered it can be seen that here there is little or no suppression of selection of the European ENTERING cards (70% versus 77% for Euro and trop respectively), and this was one of the predictions made. This prediction is upheld by the results of the post-hoc test revealing that there was a significant difference between the EATN group and the ENTA group, with the EATN group scoring a higher mean FI. However, it should be noted here that due to the nature of the indices used it is possible that these results are due to higher not-q card and lower not-p card selections in the EATN group. This alternative explanation is upheld in that the sign test performed showed no significant difference for the EATN and ENTA groups selections of p cards.

Also in this group there is the level of selections of the q and not-q card that would normally be expected on this task (18% q card and 71% not-q card). Finally, the selections of the TRANSIT cards were very low for this group (4% Euro and 11% trop). These are the results that were predicted for this group.

However, there were once again some unpredicted results and these are now discussed. For the Control group, it is not clear whether there was any suppression of the Euro ENTERING cards or not (72% Euro compared with 88% Tropical). The Cholera and No Cholera cards were also very similar in this group, which was not to be expected (36% Cholera compared with 42% No Cholera). As was mentioned in the discussion of experiment eight, here the TRANSIT cards again show a similar selection pattern to the

ENTERING cards in experiment four, suggesting that some participants are unclear as to whether those passengers who are in transit represent a threat to the rule or not.

For the ENTA group, the cholera and no cholera cards showed a reasonable level with regard to the predictions, with 53% of the no cholera cards being selected compared to 16% of the cholera cards. The Transit cards had a reasonably low selection level, 13% Euro TRANSIT, and 18% trop. However, it is the ENTERING cards here which show the most unexpected result. The European ENTERING cards, as expected, seem to be highly suppressed when compared to the other groups here, as the level of selection for these cards is only 47%, but it is in the Tropical ENTERING cards that the odd result is found, with a selection level of only 54%. It was predicted that this group would show a very high selection of these cards.

However, there was a design fault in experiments eight and nine. The likelihood is that the participants will only have examined those cards that they thought were relevant to the task, and therefore, may not have examined all the cards in the 'training set' in experiment 9. Thus those participants who ignored the European ENTERING cards, as many do, will not have gained the extra information that was on these cards, and thus will not have formed the models that it was predicted they would do, and thus the predictions made were not very strongly supported. Although experiment nine did show some evidence for the prediction that the participants were able to extract, and apply the information on the initial cards to the subsequent task.

Thus, a final experiment was designed in which the participants were required to examine all the cards in the first trial. The experimenter remained present to see if this was the case,

thus overcoming the problem of some of the participants not examining all the cards, and extracting the information that they contained.

7.9 Further Revised Computer Based Tasks.

A new experiment was designed to eliminate the flaws that have been discussed above. Here the participants were presented with 48 cards to turn over, and examine the usually "hidden" side. The participants were told that they were to examine both sides of every card, and then to click in a box below the card if they thought that the cards represented a violation of the rule or not. (Again the rule was that if the card says ENTERING on one side, then the other side of the card must include cholera among the list of vaccinations). It was these cards that contained the bias that would shape the choices that the participants would subsequently make, if they extracted the probability information that these cards contained. There were two groups used in this experiment, one was presented with cards which showed all the European ENTERING cards to be violations of the rule (no cholera vaccination). The second group showed all the Tropical ENTERING cards to be violations of the rule (again no cholera vaccinations).

The participants were then presented with a further 32 cards, and they were required to turn over those cards, and only those cards which they thought they needed to turn over to ensure that the rule was not violated. Here the cards could not be turned over to examine the other side of the cards, so the participants' selections had to be based on the side of the card that was visible to them. If the participants selected the card, then they were given the message "Card Selected for Examination". The participant could click on the card again to de-select it if they thought they had made a mistake.

By ensuring that the participants would examine every card in the initial set of 48 the original problem of participants only selecting the cards which they think are relevant is overcome.

It was predicted that the participants who were presented with the cards which showed all the European ENTERING passengers, and none of the Tropical passengers to be violations of the rule, would not show the suppression effect that has previously been observed. For the participants who were presented with cards that showed all the Tropical ENTERING passengers, and none of the European ENTERING passengers to be violations of the rule, the prediction was that the participants would show a similar pattern of selection to the original LAST experiments. Namely, a suppression of selection of the European ENTERING cards.

7.10 Experiment Ten

Method

7.10.1 Participants

The participants were 23 first year psychology students from the University of Wolverhampton, all participated on an unpaid volunteer basis.

7.10.2 Design and Materials

The participants were presented with a set of instructions regarding the first set of cards that they were to examine, these instructions are shown below:

"Please imagine that you are an immigration officer at Heathrow airport. One of your duties is to check passengers' form H. On one side of the form is the status of the passengers, whether they are entering the country, or in transit between planes. On the other side of the form is a list of vaccinations that the passengers have had. You are told by the authorities that **If the form says Entering on one side, then it must have cholera among the list of vaccinations on the other side.** You are also told that cholera is particularly prevalent in Tropical countries. On the next page is a selection of sample forms which you have been told to check by the authorities. You must click the mouse on every card regardless of the information it contains, the card will then turn over and reveal the other side. If you then believe that the passenger outlined on the card represents a violation of the rule, then you must click in the box below that card, so that an 'X' appears. Thank you, you may now begin the experiment."

After these instructions the participants were then presented with the 48 cards which contained the biases outlined above (there were three card orders used). The participants were allowed as much time as they required to look through these cards. The participants were instructed to click the mouse in the box below the cards if they thought that the cards represented a violation of the rule they had been given. (It is important to note here that the experimenter remained with the participant to ensure that they clicked every card, thus examining both sides of all the cards presented). (An example of the screens used here can be found in appendix 9).

On completion of this section of the task the participants were given a second set of instructions, as follows:

> Again, please imagine that you are an immigration officer at Heathrow airport. You must now check passengers' form H for real. Having looked through the sample forms, you know that on one side of the form is the status of the passenger, whether they are entering the country, or in transit between planes. On the other side of the form is a list of vaccinations that the passengers have had. You are told by the authorities that **If the form says Entering on one side, then it must have cholera among the list of vaccination son the other side.** You are also told that cholera is particularly prevalent in Tropical countries.
>
> On the next two pages are 32 such immigration forms which you must now check for real. In order to ensure that the rule is not violated, you must click those cards, *and only those cards,* which you think you need to check, in order to make sure that the above rule is not violated. If you make a mistake, then re-click on the card, and it will be de-selected.
>
> You may return to this instruction page at any time.
>
> Thank you, you may now begin the experiment.

The participants were then presented with 32 cards, 6 Tropical ENTERING cards, 6 European ENTERING cards, 4 Tropical TRANSIT cards, 4 European TRANSIT cards, and 6 each of Cholera and No Cholera cards. (Again 3 random card orders were used).

The participants were required to click on those cards which they believed to be a violation of the rule, the card then changed to state "Selected for Examination". If the participants

wanted to change the selection they had made then they had to click on the card again, and it would return to its original status.

7.10.3 Procedure

The participants were asked if they wanted to participate in the experiment during their seminar sessions, if they agreed then the experimenter took them to the laboratory. The participants were all tested individually in a cubicle. The participants were asked if they had mouse skills, if they did they were told to begin the experiment. If they had no mouse skills, then they were shown how to use the mouse, and then told to begin. All participants were told that if they had any difficulties then they were to ask the experimenter for clarification. Any clarification given involved a reiteration of the instructions that they had just read.

The participants were given as long as they required to finish the experiment, although most were finished in ten minutes. The participants were then thanked for their participation, and told the broad aims of the experiment.

7.11 Results and Discussion

The results for this experiment can be seen in table 7.6. These tables show that the suppression effect was reduced in the European Bias group as had been predicted. However, it was still present in the Tropical Bias group as predicted.

Percentage of Card Selections For The Participants In Experiment Ten

Table 7.6

European Bias

Entering Euro (p)	Entering Trop (p)	Transit Euro (not-p)	Transit Trop (not-p)	Cholera (q)	No Cholera (not-q)
91%	100%	0%	0%	0%	100%

N = 11

Tropical Bias,

Entering Euro (p)	Entering Trop (p)	Transit Euro (not-p)	Transit Trop (not-p)	Cholera (q)	No Cholera (not-q)
76%	96%	0%	31%	7%	92%

N = 12

It is clear from table 7.6 that the results here do support the predictions made, in that in the European bias group the suppression effect has almost disappeared (only one subject failed to select all the ENTERING cards), when making a cross-experimental comparison to experiment four. However, in the Tropical bias group the suppression effect has remained.

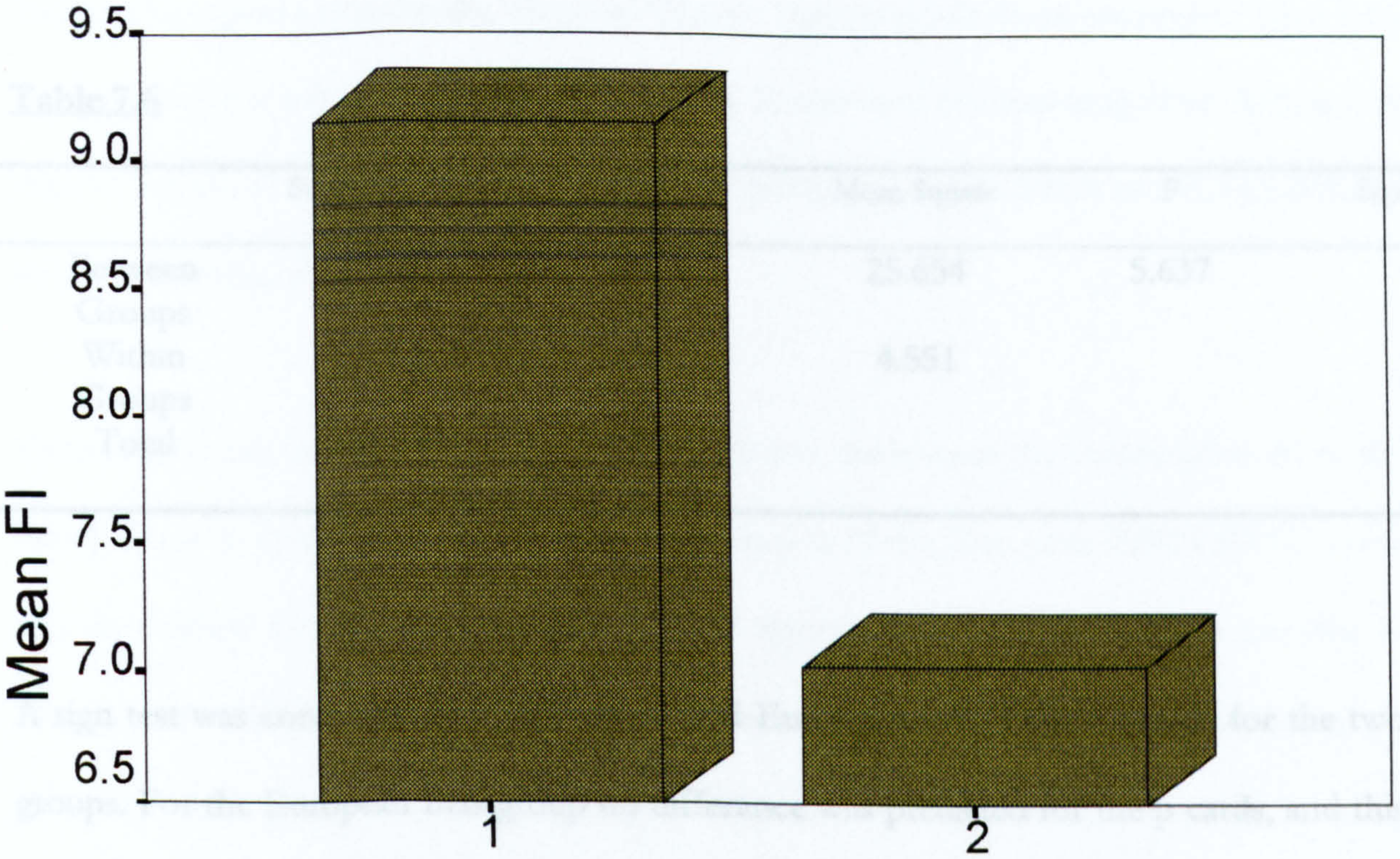
Again it was possible to calculate indices for these participants, in the same way as had previously been used (Arc Sine transforming the proportions of cards selected by each subject, and then applying the formula (p+not-q)-(not-p+q)).

The results do support the predictions that were made, with the European bias group selecting more of the logically correct cards, and not showing the suppression effect that had been previously observed.

A graph showing the mean indices for each group is give in chart 7.2 below:

CHART 7.2

MEAN FI VALUES FOR EXPERIMENT 10



GROUP

1=EATN 2=ENTA

An analysis of variance was then conducted on these indices, ($F = 5.637$, $df = 21$, $p = .027$). There was a significant difference between the selection that the two groups made in this experiment, as had been the initial prediction. From the graph, chart 7.2 and mean values it can be seen that the European bias group selected more 'correct' cards than the Tropical bias group. The results from the Anova are shown in Table 7.6 below:

Table 7.6

	Sum of Squares	df	Mean Square	F	Significance
Between Groups	25.654	1	25.654	5.637	.027
Within Groups	95.570	21	4.551		
Total	121.224	22			

A sign test was conducted on the Tropical and European ENTERING cards for the two groups. For the European Bias group no difference was predicted for the p cards, and this was observed when only 1 participant selected more Tropical p cards than European ones. For the Tropical Bias group, however, there was no significant difference between the two types of p card, with 3 participants selecting more Tropical p cards than European, no participants selecting more European than Tropical and 9 selecting the same amount of each ($p = .25$, 2 tailed). A difference was predicted here, as it would be expected that the suppression effect previously observed would be observed here also. However, there is some evidence that participants extracted the information from the cards with which they were presented, in the initial trial. The ANOVA conducted on the Arc Sine transformed data suggests that this was the case.

However, closer observation of the data reveals that this finding may, at least in part, be due to the increased selection of the Tropical not-p card in the Tropical Bias group, which would lower the FI for this group. However, the result remains that the European bias group selected more 'correct' cards than the Tropical bias group. It could be argued here that the observation of large amounts of Tropical passengers violating the rule led the Tropical bias group to select the Tropical 'Transit' card because these passengers were seen as representing a risk even though they were in Transit and not Entering. This finding can then be argued to add evidence to the idea that participants can extract and use probability information with which they are presented.

Therefore, it can be concluded that the participants did extract the information from the initial 48 cards, as the formerly observed suppression of the European ENTERING cards was diminished in the European bias group. It appears then, that participants are able to extract probability information, and base subsequent decisions on it. They appeared to do this even when this information is not explicitly stated to them, this is in contrast to utility information which needs to explicitly represented for the participants to make decisions based on it (see Manktelow and Over, 1991).

With regard to the TRANSIT cards the results are again interesting, with the only cards that were selected at all being the Tropical TRANSIT cards in the Tropical bias group. This is in keeping with some of the results from earlier experiments, in which the participants did select the Tropical TRANSIT cards. Presumably is that the participants believe that the Tropical passengers still represent a risk to keeping cholera out of the country, thus they are devising their own rationale for the rule they are given (see Cheng and Holyoak, 1985).

Finally, the Cholera and No Cholera card choices were also in accordance with predictions, that is that the participants selected almost no cholera cards, just 7% in the Tropical bias group, and almost all the No Cholera cards, 100% and 92% respectively for the European and Tropical bias groups. These selections are clearer than in the previous computer tasks, but these did contain unpredicted results in places that may have been due to the design faults mentioned above. Thus when the design was corrected then the results supported the predictions completely. Participants learnt the relevant instances, without feedback, by mere exposure to the full set of cards in the training set, once sound instructions had been given.

These results and all others will be examined in the light of the theories outlined in the introductory chapters in the General Discussion.

CHAPTER 8

8.1 General Discussion

8.1.1 Overview of General Discussion

Chapters 5, 6 and 7 provided a detailed account of the ten experiments which were conducted in this research programme designed to investigate the role of probability in reasoning with both indicative and deontic conditionals. These chapters were largely based around the empirical data produced by these experiments, and there was only sparse discussion of the theoretical implications of the findings. It is in this section that this will take place. The aim of this chapter is to provide a summary of the findings from the experiments and then to consider these findings in light of existing theoretical accounts of conditional reasoning.

Due to the nature of the research programme being split in terms of indicative and deontic conditionals, this chapter will also be broken down accordingly. Initially the results of the experiments using indicative conditionals will be summarised (experiments 1-3). The theoretical implications of these findings will then be discussed in some detail, considering three major theories that have been proposed to account for conditional reasoning. Namely, optimal data selection (Oaksford and Chater, 1994, 1995), mental logic (Braine and O'Brien, 1991; Rips, 1994), and mental models theory (Johnson-Laird, 1995; Johnson-Laird and Byrne, 1991; though see Evans, 1993).

The second section will consider the findings from experiments involving deontic conditionals (experiments 4-10). Again there will be a summary of the main findings of the experiments, and then a discussion thereof in terms of theories of deontic conditionals.

Specifically these theories will be the three theories that will have been considered as part of the discussion of results for indicative conditionals plus two domain specific theories, Pragmatic Reasoning Schemas (Cheng and Holyoak, 1985; Holyoak and Cheng, 1995), and Cosmides' Social Contract Theory (Cosmides, 1989).

The third and concluding section of this chapter will provide a summary of this theoretical debate and provide discussion of required research, and an overall summary of this research programme itself.

8.2. General Summary of Findings Using Indicative Conditionals

Initially, an experiment was run in which the participants were given an indicative conditional in a horse racing scenario, which contained both probability and utility manipulations. As was noted earlier it is extremely difficult to separate these two constructs in thematic experimental manipulations, as participants will always assess epistemic utility, that is the pay-off associated with actions aimed at fulfilling epistemic goals (see Levi, 1984 for a philosophical view of epistemic utility). For example, mothers will always prefer tidy rooms (Manktelow & Over, 1991). The results from this experiment did suggest that there was some way in which participants take account of probabilistic factors when making choices on the selection task, and that the scenario provided in this experiment was aiding participants' selection on this version of the task. Specifically, there was evidence of increased selection of the not-q card, the card typically ignored when using indicative conditionals and the selection task, except in the therapy experiments (Wason, 1968) (see Evans, Newstead and Byrne, 1993 Chapter 4, for a detailed review of the selection task literature).

However, the initial results were not as clear as had been predicted, and so revisions were made to the format of the task to make it more ecologically valid, to gain a better understanding of human reasoning with indicative conditionals which contain some probabilistic factor. This resulted in the introduction of the Large Array Selection Task (LAST), which was designed to make the task more like those which participants would have encountered before in a real world-reasoning situation. When participants make probability judgements in an experimental situation they seem to be better equipped to consider those problems which are presented within a frequentist framework (Gigerenzer, 1991 1996, but see also Kahneman and Tversky, 1996) i.e. considering relative frequencies of items, rather than single-event probabilities expressed as proportions.

Therefore, it was felt that the experiments requiring the participants to make probability judgements should take account of this, and experiment 2 as presented in 5.6 was the first to do so. Thus the LAST was introduced in order to include the probability manipulations that had been inserted into the task. The LAST was necessary here as it enables a probability manipulation to be inserted in the task, within an ecologically valid framework. The LAST also permits a large amount of data to be collected without running large participant numbers. Finally it allows a more sensitive measure, which is of particular relevance to probabilistic manipulations. The LAST allows participants to display their probability judgements in a way that they could not do in a standard four-card task. It allows a participant to select, for example, *all* the examples of a card that they believe to be of particular relevance, but only *half* the examples of a card they consider to be of less importance to the task. It is interesting to note here that the LAST is similar to the Reduced Array Selection Task (RAST, Johnson-Laird and Wason, 1970, Wason and Green, 1984), in

that they are both manipulations of the selection task which allow participants to make successive selections of cards on the one task.

When the LAST was employed the participants' choices of the not-q card (the card which participants typically ignore on indicative tasks, see Chapter Two for details), rose as high as 51%. This suggested that the introduction of probabilistic information into the indicative task did lead to an increase in the selection of these cards. However, it was still not to the levels that had originally been predicted: that is to the level of selection of the p and q cards on the same task, which were at 86% and 78%, respectively, (averaged across tasks in experiments 1 and 2). However, it should be noted that 50% is a high level of selection for a version of the selection task involving indicative conditionals such as the version used here (see Dominowski, 1995).

Finally, a revised version of the gain scenario from experiment 2, which had produced the highest selection of the not-q card in the LAST experiment, was designed. The aim here was to attempt to raise the selection of the not-q cards to the same level as the p and q cards on the same task. This involved labeling the betting slips as week 1, week 2 etc. so the participants were clear on what the cards represented, as well as changing the not-q cards to 'LOST' rather than 'DID NOT WIN'. However, these revisions did not lead to facilitation on the task, as the participants still chose the p and q cards more than they did the not-q. Indeed the most interesting result from this version of the task was that there was a drop in the selections of the p and q cards to 64% and 73% respectively, with selections of the not-q card at 46%. These results are difficult to reconcile with any of the theories that have been examined, as there is no explanation of any version of the task where the p card should not be the most commonly selected card.

This represents an overview of the results from the three experiments that used indicative conditionals rather than deontic conditionals. The basic findings were that although there was an increase in the selection of the not-q card, when compared to previous indicative task performance with thematic materials (see Manktelow and Evans, 1979), it did not rise to the same level as the selections of the p and q cards. The prediction was that participants should select the p, q and not-q cards at the same level. The following section will consider how the theories would account for these findings. The theories to be considered are optimal data selection (Oaksford and Chater, 1994, 1996; Oaksford, Chater, Grainger and Larkin, 1996), mental logic, (e.g. Rips, 1994; Braine and O'Brien, 1991) and mental models theory (Johnson-Laird, 1995; Johnson-Laird and Byrne, 1991).

8.3.Theoretical Discussion and Implications of Findings

8.4 Optimal Data Selection

Firstly, the Optimal Data Selection theory of Oaksford and Chater (1994). This approach is of particular interest here as its basis is probability theory, and therefore is of particular relevance. Oaksford and Chater claim that individuals do not view the selection task as a reasoning task at all, but rather it is seen as a hypothesis testing exercise. When the participants are given the rule, or conditional, they are required to assess whether the rule is true or false. Reasoning researchers have traditionally argued that participants should look to falsify this rule, and thus they should select the p and not-q cards. However, Oaksford and Chater argue that this falsificationist view of science, as proposed by Popper (1959), is now outdated, and is no longer the predominant view in the philosophy of science.

Consequently, it is unreasonable for reasoning researchers to require their participants to conform to this as the ideal performance that everybody should strive to achieve.

Oaksford and Chater state that when the task is viewed as a hypothesis testing task participants will select those cards which are most likely to decrease their uncertainty about the rule, and thus select those cards which contain the most information. Oaksford and Chater propose that these cards are as follows:

p>q>not-q>not-p

And that this should be the pattern of results that participants should display. The literature on indicative conditionals, as discussed in Chapter 2 largely displays this pattern of performance (see also Evans, Newstead and Byrne, 1993, Chapter 4).

With the indicative conditionals that have been used here this was indeed the predominant pattern of results throughout these experiments. Only two of the experiments failed to conform to the predicted pattern of results from Oaksford and Chater and these were experiment 1b where the pattern was: p=q>not-q>not-p; and experiment 3a where the pattern was q>p>not-q>not-p. The results from experiment 3 were noted as being strange in the discussion for that experiment, 5.12.

Therefore the theory proposed by Oaksford and Chater does seem to account well for the majority of the data presented here. There are a number of criticisms which can be made of the ODS theory on various levels. The theory does seem to have the failing of explaining what is happening without explaining how. To put this more simply the theory is a

computational theory, rather than a psychological one. Oaksford and Chater presented equations that they claim can account for the patterns of results that have been found in the majority of the selection task literature. These equations are based on the idea that as participants are testing a hypothesis, their goal is to reduce their uncertainty. Oaksford and Chater claimed that reduction in uncertainty must equate with information gain (though see Evans and Over, 1996a for a criticism of this equality). However, they offer no account of how this theory is actually put into practice by people who are engaged in reasoning. Whereas mental models and mental logic have made some attempt, albeit incomplete, to account for the psychological processes taking place.

Thus it appears, to use Marr's (1982) terminology that Oaksford and Chater have provided the computational level of their theory there is no algorithmic level of *how* people actually make the selections they do when presented with the selection task in its indicative form. It can be claimed, on these grounds then, that ODS theory is not a psychological theory, and that for the theory to truly offer an account of human reasoning processes it must be developed to provide a 'HOW' as well as a 'WHAT'. In addition to this, it seems to be an odd claim of the theory that participants are applying a Bayesian analysis to the selection task, when there is an increasing amount of evidence to show that people regularly fail to conform to Bayes' theorem in experimental tasks (Sutherland, 1998). However, it is important that the predictions that Oaksford and Chater make are examined here in detail and discussed in light of the findings of the current research.

One of the major assumptions of Oaksford and Chater's theory is that of rarity, and this assumption has implications for the ordering of card selections that are predicted, this assumption is not without criticism (Laming, 1996). Thus rarity will be examined here in

detail, as will the applicability of the ODS model to the findings of the indicative experiments.

Oaksford and Chater, (1994, 1996) calculated scaled expected information gain (SE(Ig)) for the four cards in the selection task based on the assumption that the occurrences described in p and q are rare (in this case, tipping and horse, and horses winning races). Oaksford, Chater, Grainger and Larkin (1996) illustrate rarity using the following example:

> "All ravens are black, the probability that any given bird is a raven is low, as is the probability that it is black."

They argue that this holds true for the vast majority of categories that are used to construct hypotheses about the world.

Thus, what Oaksford and Chater would claim here is that in the rule *'If Racing Weekly tips a horse then it wins'* both the tipping of horses and horses winning races is rare. However, it is argued here that in a version of the selection task such as this where participants are given a lengthy scenario and rationale for the rule that they have been given the rarity of p and q begins to decrease. The rationale must provide participants with enough information to discard the rarity assumption. Specifically, their thinking has been directed to the issue in the task and thus the irrelevant factors of thought which allow for the rarity assumption to hold have been violated (e.g. that it could be any bird), thus negating the need for the assumption within the ODS framework. In addition to this Racing Weekly is there *to* tip horses, and thus rarity is uncertain here, in that it is not rare for Racing Weekly to tip horses. Thus the rarity assumption is challenged for the p card. With regard to q, however,

rarity holds - horses that win must be more rare than horses that lose, because there are more than two horses in each race. In this version of the selection task it is argued that rarity holds for the q card, but is uncertain for the p card. Rarity for the p card is that fewer horses are tipped (p) than not-tipped (not-p). While it is likely that rarity occurs here it is uncertain for participants, and this would have an effect on their selections, because the rarity assumption is challenged. Finally, it is difficult for Oaksford and Chater to account for the high selections of the not-q card, because, according to their account, this should only occur when rarity is challenged, and this is not in the case for the q cards, as argued above.

In summary then Oaksford and Chater's theory provides a reasonable fit for the data considered thus far. However, the questioning of the rarity assumption suggests problems with the theory as it stands. Given the good fit for most of the data, however, it is unclear whether the theory can fully account for these findings or not. However, it remains that the viability of a theory formed from Bayes' theorem must be questioned given Gigerenzer's compelling account of the importance of a frequentist framework for probabilistic experiments.

8.5 Mental Logic

As was discussed in Chapter Two mental logic approaches advocate some form of content-independent rules that are applied by humans when they engage in reasoning. However, O'Brien (1993) has claimed that mental logic has been misrepresented in that it has been argued that a theory that adopts this approach would be unable to account for the content effects which occur with versions of the selection task which clearly indicated that content

effects are present. O'Brien argues that the mental logic approach makes no such claim, but rather argues for some kind of pragmatic principles, such as Pragmatic Reasoning Schemas which are evoked to aid reasoning at the appropriate times.

Presumably then, when participants are presented with a conditional of the form that was used in experiments 1-3 here, some form of pragmatic principle would be used by the participants to solve the task, rather than applying the abstract formal rules. Mental logic proposes that humans use abstract rules to reason when the form of the problem with which they are faced does not evoke the pragmatic principles.

It is difficult to understand what mental logic would suggest as a prediction. As was discussed in Chapter 5 the standard p and not-q prediction is not made. It is important to the reasoner to look at the WON (q) card as well as to look at the not-q card. There is nothing that exists within the current formulation of mental logic that can capture this need to examine the q card as well as the not-q card. A mental logic approach then provides very little of any use when trying to assess the likelihood of q given p, as in this task. However, it remains the case that it is predicted that there should be an increase in the selections of the not-q card, and Braine and O'Brien's approach does state something about this.

8.5.1 Braine and O'Brien's Model

O'Brien (1993) refers to the mental logic theory of Braine and O'Brien (1991) as The Model, and the same designation will be adopted here. O'Brien (1993, 1995) claims that the inferences required to complete the selection task are beyond the scope of the core and feeder schemas of The Model, and as such it should be no surprise that participants fail to

solve the task correctly. Thus it seems that The Model would predict that there should be no increase in the selections of not-q cards here, as such reasoning is outside of the inference schemas within The Model.

However, O'Brien (1993) also noted that there exist algebraic versions of the selection task which participants correctly solve e.g. Griggs, 1989 (algebraic versions of the task are typically the most difficult for participants). He claims that such patterns of performance "are difficult to explain unless one assumes some logical reasoning" (O'Brien, 1993, p.123). This claim seems difficult to reconcile with O'Brien's concurrent claim that the selection task lies outside the realm of The Model. Either The Model *does* provide the inferences necessary to solve the selection task or it does *not*. It should not be claimed that it does provide the inferences when participants do solve the task correctly, but does not provide them when participants fail to demonstrate logical responses, which is what O'Brien's claim appears to be.

What would the prediction made by The Model be for the indicative rule experiments presented here? Basically there are two possible predictions which could be made from The Model. If there is a significant facilitation effect due to content on this version of the task then it should invoke the pragmatic principles component of the theory. However, this is unlikely as O'Brien cites Pragmatic Reasoning Schemas (Cheng and Holyoak, 1985, Holyoak and Cheng, 1995) as a suitable theory to handle such cases. This theory is domain specific to deontic reasoning, and as such does not apply to indicative conditionals of the form presented to participants here. Thus it is necessary to consider the predictions that could be made from The Model without consideration of the pragmatic principles component. This leaves the theory in its more abstract form, which considers the selection

task to be outside the range of the inference schema provided within The Model. O'Brien (1993) suggests that the reasoning required to turn even the p card is considerable, and is not just the MP inference often suggested (e.g. Cheng and Holyoak, 1985).

It is important to consider the reasoning that O'Brien (1993) suggests would be required for participants to select the predicted cards as presented in this research programme. It involves initially supposing that the rule ("If Racing Weekly Tips a Horse then it Wins") is true. Given this supposition and a card showing "Tipped by Racing Weekly" application of modus ponens should lead the participant to the conclusion that the reverse side of the card should show "Won". However, O'Brien claims that this is insufficient reason for the participant to select the p card. To select the card the reasoner must now consider the possibility that the reverse side of the card shows "Did Not Win". This would then allow the supposition that the rule is true to be falsified by reductio ad absurdum, because there is the possibility that there could be "Did Not Win" where there should be "Won". O'Brien would therefore suggest that the selection of the p card was outside the realm of the inference schema of The Model. Selection of the not-q card (in this case "Did Not Win") is an even harder inference to make than that required to select the p card as described above. It appears then that The Model would predict low selection rates for both p and not-q cards, such as participants resorting to matching.

The results from the indicative experiments failed to uphold either of these predictions derived from The Model of Braine and O'Brien (1991). Thus it seems that Braine and O'Brien's theory is unable to account for the findings of experiments 1-3 in the present research programme. However, Braine and O'Brien's theory is not the only Mental Logic based theory that has been proposed to account for performance on the selection task. Rips

(1983, 1994) has also presented an abstract rule based account of the selection task. This will now be considered here.

8.5.2 Rips' Account

Rips (1994) presented a compelling and lucid account from the mental logic perspective, and attempts to explain how such a position might account for the diverse patterns of performance observed in participants when presented with the selection task, in its various forms. Rips presented his account of the selection task in the following way. Say the rule was of the form "If there is a vowel then there is an even number". The four cards used in this task would be, for example, E K 4 and 7. The cards can then be combined with the rule to produce the following sets, from which participants must generate conclusions in order to complete the task successfully.

1. E card IF vowel THEN even

   Vowel.

   ?

2. 4 card IF vowel THEN even

   Even.

   ?

3. K card IF vowel THEN even

   Not Vowel.

   ?

4. 7 card IF vowel THEN even

NOT EVEN

?

Rips presented his account using the PSYCOP model. When presented with the selection task Rips' model would select the E card only, as is the response of approximately one third of participants. The reason for this response from PSYCOP is that the only set for which PSYCOP can produce an inference is 1 above. For all the other sets PSYCOP draws a blank, and fails to produce an inference. The important card in relation to the selection task is the 7 card (not-q card). PSYCOP produces nothing for this card, as this would require possession of a forward modus tollens rule, which the PSYCOP model does not have. No backward search is triggered either due to the lack of a subgoal. Thus PSYCOP would select only the E card – as around 30% of participants do.

However, how would the model explain the modal response of E and 4? Rips accounts for this by employing the assumption that the converse of the rule is also assumed to be true "If even then vowel". When given this rule PSYCOP selects both the E and the 4 cards. Rips' model can account for the findings that the E and 4 cards are selected, and that the E card alone is selected, but how does Rips account for correct task performance? Rips argues, in the same way as O'Brien (1993) that correct solution of the task requires participants to project possible values on to the reverse sides of the cards. This process is effortful, and as a result, rare, thus accounting for the correct solution rate of c. 10%. As with other theories in this section the issue here is: what would Rips' model predict for the present experiments, and can it account for the findings?

Rips appears to advocate some form of memory cueing hypothesis in order to explain content effects on the selection task, he claims that the information received from memory would enable participants to draw inferences that they cannot when the task is presented in its abstract form. Although Rips discusses Cheng and Holyoak's PRS approach to the selection task, this is of little relevance here as PRS theory only attempts to explain performance on deontic versions of the task concerned with permission, obligation etc. Given that here participants are producing a reasonably high level of not-q card selections Rips would presumably explain the results in terms of memory cueing. The prediction would be the that if the participants have enough experience of the subject matter used in the task they should respond accordingly i.e. by selecting the not-q card.

Rips acknowledges that a full account would need to be given of the retrieval process involved in memory cueing, but nonetheless adopts the framework as workable. However, although there is some evidence for memory cueing (Manktelow and Evans, 1979) the process is largely unexplained. For example, in the experiments here, how much knowledge of the area would a participant need? In addition to this if a participant does employ some form of memorial cue, then presumably they are no longer employing a purely rule based strategy, and thus Rips' model would fail to apply in a situation such as this. Given that the participants are not then reasoning via rules they must be using some other process by which they can produce the predicted response pattern and Rips' model makes no mention of a process such as this. In summary then, the theory remains unable to account for the pattern of performance demonstrated here.

Two approaches from the view of mental logic have been considered in detail here. Both the theories (Braine and O'Brien, 1991 and Rips, 1994) have interesting points to make regarding participants' performance on the selection task, due, in part, to the suggestion that abstract rule theories are unable to explain the effect of content on the task (Cheng and Holyoak, 1985). However, neither of the theories is able to make predictions in line with the findings or account for the results observed in the present experiments.

8.6 Mental Models Theory

The remaining theory is mental models, in a formulation of some sort (Johnson-Laird and Byrne, 1991; but see calls for revisions to the theory in Manktelow and Over, 1992, Evans, 1993, Evans and Over, 1996b, Hardman, 1996). Mental models theory was discussed in some detail in Chapter Two, and some form of revised mental models theory is favoured here.

Johnson-Laird and Byrne have said that any form of the selection task that leads to the fleshing out of further models will lead to increased selection of the not-q card, typically ignored by participants when presented with the standard task. Therefore, what models will the participants construct here? Why should this version of the task lead to the selection of the not-q card rising as high as 50%, when the q card selections do not diminish, as is normally the case when not-q selection rise? (See Dominowski, 1995 and Evans, Newstead and Byrne, 1993).

Johnson-Laird and Byrne (1991) claimed that when presented with the standard version of the selection task the participants construct the following initial models, given that the rule

is of the form If p then q. Note here that the 1991 notation is maintained although Johnson-Laird has recently revised the format (Johnson-Laird, 1995).

[p] q

...

where '...' represents implicit models that are yet to be fleshed out, and '[ ]' represents that an item is exhaustively represented. Johnson-Laird and Byrne claim that as much information as possible will be represented implicitly due to limitations of working memory, and other cognitive constraints.

When the models are fully fleshed out then they are as follows, because p cannot appear in any model except with q according to the initial model set (as represented by '[ ]' above):

p q

¬p q

¬p ¬q

There is also the issue here of the biconditional interpretation of the rule: If and only if p then q. When this interpretation is given to the rule the models constructed are obviously different to those described above. However, such models are not of great significance here as it is unlikely that the rule used in these experiments would be subjected to a biconditional interpretation by participants. Specifically, it is argued here that it is unlikely that participants would interpret the rule "*If Racing Weekly Tips a Horse Then it Wins*" as *"If and only if Racing Weekly Tips a Horse Then It Wins"*. Such an interpretation would be a violation

of the knowledge of the participants – that merely tipping a horse does not guarantee its success, and if this was the case the whole exercise of tipping horses and placing bets would be rendered futile. There is substantial evidence that participants use existing knowledge when approaching reasoning tasks (see for example belief bias in syllogistic reasoning, where participants rely on their own beliefs rather than their reasoning processes, Newstead, Pollard, Evans & Allen, 1992).

When we consider these models in the sense of the rule that the participants here were given we have the following models:

tipped	won
not tipped	won
not tipped	lost

This fleshing out of the models, claim Johnson-Laird and Byrne, leads participants to consider the not-q card, which the initial representations do not, and hence this leads to the selection of the not-q card, which was shown in the results here. Thus it is possible to make the prediction that a scenario such as that used here is one that would lead to fleshing out of initial models. This fleshing out leads to an explicit representation of the p, q and not-q cards in the models, and these cards were then selected by participants. Those participants who failed to select the not-q card had not fleshed out the models fully, and subsequently selected only the p and q cards. It is therefore, possible to account for the patterns of selections shown by participants on this task via the mental models theory.

However, there is an issue that remains unsolved here with respect to mental models theory. This is the question of focussing. Why should participants here flesh out their models and show evidence of focussing? Legrenzi, Girotto, and Johnson-Laird (1993) both explained and examined the role of focussing in deduction generally, and on the selection task, which is of most interest here. Legrenzi et al. argue two points in relation to focussing and the selection task:

> "1. Participants consider only those cards that are explicitly represented in their models of the rule, that is, they focus on these cards.
> 2. They select from these cards those for which the hidden value could have a bearing on the rule, that is, those that are represented exhaustively (with square brackets in our diagrams of models)."
>
> Legrenzi, Girotto and Johnson-Laird, 1993 p.49.

The question to address here is why should participants be fleshing out and not focussing? The answer to this lies in a closer examination of responding. If the task was purely one that led participants to focus on the not-q card, then there would be a high level of not-q responses, as observed, but it is the other patterns of responses that give the answer to the question above. Because some participants are showing a selection pattern of p, q and not-q then the task is resulting in fleshing out for these participants, hence they are selecting the cards represented in their models. The task does not lead to focussing on the not-q card, because selections of the q card remain higher than those of the not-q card. It is therefore argued here that the task is leading to fleshing out of initial models, and thus selection of the not-q card increases, rather than a qualitative shift to consideration of the not-q card as opposed to the q card, as focussing would do. The point is that those participants who

represent the not-q card in their model will see it as important, and those who stick with their initial representation will not. There is no manipulation of the task *per se* here that should give rise to a focussing effect. As Legrenzi et al note there will always be individual differences in the selections made by participants – some correctly solve the abstract task, and some fail to solve deontic versions.

8.7 Summary of Discussion of Indicative Conditionals

Thus it is argued here that the first set of results (those from the indicative conditionals) favour a mental models explanation than any other. Mental models appears to be the only proposed theory which can offer any account of participants' ability to select the not-q card in an indicative task such as this, while still allowing for the high selection rates of the p card as predicted. Neither Oaksford and Chater' optimal data selection theory, or any formulation of mental models has the flexibility to account for the findings reported in Chapter 5. Attention will now be turned to the deontic conditionals, and the various theories that have been offered to account for reasoning with conditionals in the deontic domain.

8.8 Summary of Findings Using Deontic Conditionals

Experiment four was the first of the deontic selection tasks to employ the LAST methodology. This experiment involved the use of a classic task from the literature on deontic reasoning, namely Cheng and Holyoak's (1985) immigration task. The changes that were made to the original task were to present the task in the LAST format, and to include probability information on the cards with which the participants were presented. In addition to the

information that the participants were given in Cheng and Holyoak's original task the cards now contained information about the passengers' country of origin on the p cards, and these fell into two categories, either Tropical or European.

It was predicted that the participants would see those passengers who were ENTERING the country as representing a greater risk if they were of Tropical origin than those passengers who were of European origin. Even though, of course, any passenger who was a violation of the rule that the participants had been given logically represented a risk regardless of their country of origin.

The results from this experiment did indeed support the predictions that had been made: the participants selected the Tropical p cards at a far higher level of selection than the European p cards, 98% for the Tropical p cards, and 57% for the European p cards.

Thus having found that probability elements did affect the choices which participants made on the task, it was essential to establish whether these changes were due to a facilitation of selection of the Tropical ENTERING cards or due to a suppression of the European ENTERING cards.

Therefore, a second experiment was run in which the LAST was used, but the probability information was removed from the cards, so as to examine the problem outlined above as to whether the effect of probability information was facilitation or suppression.

There were two conditions used here, one in which the participants were given a rationale with their instructions, whereas in a second condition no was rationale was given. The results from

this experiment revealed that the effect of the probability information on the cards was to suppress the selection of the European ENTERING cards, rather than to facilitate the selection of the Tropical ENTERING cards. In the rationale conditions the frequency of selection of the p card was 80%, and in the no rationale condition the frequency of selection of the p card was 98%, therefore, collapsing over the two conditions, the p card was selected 89%.

The effect of the rationale sentence was examined briefly above, but it was felt that the role of the rationale should be examined clearly. Thus a version of the task was designed in which the rationale sentence was omitted but the probability information on the cards was reinstated (that is the information about the country of origin of the passengers was on the p and not-p cards). The results from experiment six could not be clearer, when the rationale sentence was omitted the ENTERING cards were selected at the same level, regardless of the country of origin: the Tropical and European p cards were both selected 93% of the time by the participants.

The results from this experiment suggest that probability is similar to utility in its role in deontic reasoning (see Manktelow and Over, 1991, for a discussion of the role of utility in deontic reasoning). More specifically, probability needs to be explicitly represented in the rationale or scenario of the task, or be familiar in participants' experience for participants to react to such manipulations. This may only be the case in laboratory based tasks, however. The participants do not explicitly represent utility and probability when they are not going to be really affected by the outcome of the choices they make; thus these elements of a task need to be made explicit for the participants to make their selections based on this information. However, it is difficult to imagine that humans do not examine their own 'real-life' decisions

sufficiently to bring utility and probability elements into the equation, even though such factors may not be explicit even in these real world tasks. Indeed it is clear, merely from the survival of the human race that people are able to make decisions, often quickly, that result in survival over non-survival, for example when driving. Ball, Sutherland and Quayle (1997) noted that although participants may be motivated to behave in a logical manner, they fail to do so and thus rely on heuristics, or fail to fully flesh out their models. It is possible that this tendency is increased in an artificial environment.

Experiment seven was run to assess the role of probability when the probability information was on the consequent cards in the selection task, rather than on the antecedent cards as was the case in experiments four to six.

The participants were presented with instructions which were similar to those in the previous experiments, except the instructions told the participants that one side of the card showed whether the passenger was entering the country or not. Whereas the other side of the card showed both the diseases against which the passenger had been vaccinated, along with the country of origin of the passenger. Thus it was predicted that the participants would select those not-q cards which showed that the passengers were from Tropical countries, rather than those which showed that the passengers had come from European countries. Again the results were consistent with the predictions which were made. The participants showed a bias towards selecting those q cards that were of passengers from Tropical countries (Tropical q card selections were at 67%, against 40% for European q cards). As was stated earlier, this is a quite remarkable result considering that the participants could see the vaccinations which the passengers had had, as well as the country of origin, and so there really should have been no bias shown in this experiment. Although no experiment was run to determine whether the

effect was one of facilitation or suppression, it is not unreasonable to assume that it is again a suppression effect, as was the case in experiment four.

Finally, computer based experiments were run that would assess whether participants were able to extract information about the likelihood of the rule being violated (experiments 8 - 10). Experiments 8 and 9 contained design errors, consequently only the results from experiment 10 will be considered here. Experiment 10 clearly demonstrated that participants were able to extract probabilistic information when presented with cards which displayed a bias towards Tropical passengers being a violation of the rule. These data are in keeping with Gigerenzer's claims that humans base probability decisions on frequencies. Participants were presented with 48 cards that contained information about violations. The information which participants received at this time clearly influenced any subsequent choices they made on the task itself. Specifically the suppression of the European selections previously observed disappeared when the initial 48 cards showed that the European passengers violated the rule, and the Tropical ones did not.

Consolidating the results from the initial indicative experiments, the deontic based tasks showed strong evidence for participants being aware of probability when making selections on the task. There was suppression of selecting those cards which showed that the passenger was of European origin, when the implications regarding Tropical countries and the likelihood of carrying cholera were explicitly presented to the participants.

8.9 Theoretical Discussion and Implications of Findings

8.9.1 Section Overview

The results from the deontic experiments will now be discussed in terms of the major theories that have been proposed to account for performance on deontic selection tasks. Firstly, the domain-specific theories Pragmatic Reasoning Schemas (Cheng and Holyoak, 1985 and Holyoak and Cheng, 1995); and Social Contract Theory (Cosmides, 1989). Secondly, other theories of reasoning will be considered: Optimal Data Selection (Oaksford and Chater, 1994, 1996); the mental logic approach (O'Brien, 1993, Rips, 1994); and the mental models approach (Johnson-Laird, 1983, Johnson-Laird and Byrne, 1991).

8.10 Pragmatic Reasoning Schema Theory

Firstly, Pragmatic Reasoning Schema Theory, as discussed in Chapter Three. Cheng and Holyoak (1985) contend that once the rule in the task can be used to access one of the production rules in the schema then the participants should go on to solve the task successfully. However, in the tasks concerned with deontic reasoning above, experiments 4-10, if the participants have accessed the production rules, then why do they show the biases which they do?

As it stands the theory of Pragmatic Reasoning Schemas (PRS) would suggest that if the production rule has been accessed then the participant should select the corresponding cards, usually the p and not-q cards. Therefore, given a rule in the form "*If a person has ENTERING on one side of their immigration card then they must have CHOLERA on the reverse side*" PRS theory

would predict that participants would select the p and the not-q cards. In this version of the task this would be the ENTERING card, and the card list of vaccinations that did not include cholera. The findings of the deontic experiments present a problem for PRS, there is no suggestion here that the rule should be 'blocked' in certain cases. However, that would seem to be what is happening here. The selection of the European p cards has been suppressed, thus participants are failing to make the modus ponens deduction required. Byrne has presented an account of the blocking of valid modus ponens inference (Byrne, 1989). From the work of Cummins, Lubart, Alksnis and Rist (1991) it is possible to suggest that the information that a passenger has come from a Tropical country acts as some kind of 'disabling condition' blocking the necessary inference.

The difficulty of probability for PRS theory is comparable with the problem that the role of utility presents the theory, and as has been argued earlier the two constructs of probability and utility are closely related. The efforts that Holyoak and Cheng (1995 a and b) have made to ensure the survival of pragmatic reasoning schemas in light of the evidence that utility has an effect on reasoning have decreased the parsimony of the theory considerably. This was initially one of its great strengths, and this has now been reduced. Holyoak and Cheng invoke the idea that rights and duties are complimentary, and thus, using Manktelow and Over's example, the son has a duty to tidy his room to give him the right to go out and play. Holyoak and Cheng (1995a) claim that it is this complimentarity which leads to the perspective effects on the selection task which have been observed. However, this explanation has been seen as being largely unsatisfactory by a number of commentators (Johnson-Laird and Byrne, 1995; Oaksford and Chater, 1995; Over and Manktelow, 1995, but see also Holyoak and Cheng, 1995b). If the theory were now, as appears to be the case, required to be extended again to account for probability it would become unparsimonious, and overly complicated and

convoluted. Oaksford and Chater (1995) criticised Holyoak and Cheng's (1995a) formulation of the theory for being unparsimonious. The addition of further components to account for the effect of probabilistic information would increase this difficulty with the theory. In summary then the current version of PRS theory would need to provide some mechanism to selectively block selections of the p card, which is what happened in the experiments presented here. Neither can the theory explain the selective blocking of not-q card selections as observed in experiment 7.

8.11 Social Contract Theory

The second of the domain specific theories to be discussed is that of Social Contract Theory, (SCT) as presented by Cosmides (1989). Cosmides has argued that humans are evolutionarily pre-programmed to avoid cheaters, those who take a benefit but do not pay a cost. The theory would suggest that any selection task that fits this cost-benefit framework should yield facilitated performance, in terms of an increase in p and not-q responses. It can be seen that it is possible to view the immigration rule in these terms. It is a benefit to enter a country, but there is the cost of having inoculations, and Cosmides would argue that this is the reason the task produces such a high level of facilitation (c. 90% p and not-q responses - Cheng and Holyoak, 1985). In the same way as PRS theory it is difficult to see any reason how the theory could account for the drop in p card selections when they are from a European country, as was observed here. Thus the rule is in accordance with the cost-benefit structure, and yet it fails to present the predicted pattern of behaviour.

In summary, SCT is unable to account for the findings of the research here with deontic conditionals, because the observed pattern of performance is not in accordance with the

predictions of the theory. PRS theory is also unable to account for the findings for the same reason. It is suggested here that domain specific theories are too limited, in that they suggest that once a rule is of a certain type then facilitation *must* follow, but this is not strictly true. There are so many factors which are capable of influencing reasoning processes, utility, probability, explicit negatives, instructional factors, (see Yachanin and Tweney, 1982; Jackson and Griggs, 1990; Manktelow and Over, 1991; Kirby, 1994). Domain specific theories seem largely to ignore these and advocate some form of irreversible reasoning route once the rule fits the schema. It is argued here that a theory of this nature ignores too many factors to hope to offer a true, justifiable account of the richness of human reasoning.

8.12 Non-Domain Specific Reasoning Theories

Having considered those theories proposed to account only for deontic reasoning, attention will now be turned to non-domain specific theories of reasoning. These theories have been considered in the section on reasoning with indicative conditionals (8.3 – 8.7). However, they propose alternative accounts of reasoning with deontic statements and thus will now be considered again in light of the findings from experiments 4-10.

8.13 Mental Logic

The third class of theory that will now be considered is that of mental logic. As was discussed in section 8.4 mental logic theory claims a form of abstract reasoning rules in the mind that we use when we engage in the reasoning process. However, Braine and O'Brien argue that when we are presented with a thematic version of the selection task we use some form of pragmatic principle rather than abstract rules. Mental logic and abstract rules are similar in that they both require some component of the theory that specifies when the rules of reasoning will be

applied *selectively*. Given the nature of the task used here, it is with this pragmatic principle that this section must concern itself. Braine and O'Brien (1991) suggest that the permission schema of Cheng and Holyoak's (1985) PRS theory is a suitable candidate for how the content of propositions affects the way in which they are constructed. However, the limitations of this theory have already been discussed in section 8.8.2 and there is no need to repeat them here. Suffice to say that the criticisms raised in that earlier section render Braine and O'Brien's approach to reasoning performance as incomplete and unsatisfactory.

Rips (1994) has also presented an account of reasoning from a mental logic or abstract rules perspective. However, Rips also advocates pragmatic principles, and, as Braine and O'Brien did, advocates Cheng and Holyoak's (1985) PRS theory. This theory has been demonstrated to be inadequate to account for the findings here in section 8.8.2

In summary then of the mental logic approach, these theories have been criticised for the lack of pragmatic factors that they consider. The results that are presented here also give mental logicians a difficulty, in that theories of mental logic fail to give any account of probabilistic factors. Specifically, there is no account of why the rules should be applied to some of the cards, and not to others if it is true that the rules are abstract. However, it is the case that the participants here did selectively choose cards: the Tropical ENTERING cards far more than the European ENTERING cards, even though they have logically the same value.

In addition to this, why in the initial indicative versions (horse race scenario) of the selection task should the participants' selection of the q and not-q cards be so high? Again this suggests that participants are failing to follow logic when they are presented with problems which contain probabilistic factors. Presumably they are applying some other method of reasoning to

the tasks, which allows them to take account of the probability information which they are being given, which is not the case with mental logic.

8.14 Optimal Data Selection

Oaksford and Chater (1994) have offered an account of selection task performance based on Anderson's theory of rational analysis (Anderson, 1990). It is similar to their account of abstract task performance, in that it is based on information gain. However, the assumptions made to account for abstract task performance are altered, and it is these changes which allow the theory to account for performance observed on the deontic task (see section 8.3.1).

As was discussed earlier Oaksford and Chater claim that the selection task is a form of hypothesis testing, or certainly is treated as such by participants when they are presented with the task. This is why they claim their theory can be used as an account for both the abstract and deontic selection tasks. The change that they make here is to claim that rarity is overturned in rules in a deontic form, and this, in turn, affects the predicted card orders. In a deontic framework Oaksford and Chater's model predicts that the following card orderings should be observed based on the potential information gain of the cards:

p>not-q>q>not-p

These card orders are typically observed in deontic versions of the selection task (see Dominowski, 1995 for a review of the effect of content on reasoning). What is important here is how do the predicted card orderings from Oaksford and Chater's model compare with the observed selection patterns? Oaksford and Chater (1996) have claimed that the findings from

experiments 4-7 (as published in Manktelow, Sutherland and Over, 1995), offer support for their Rational Analysis model, but the extent of this support will be examined here.

As Oaksford and Chater make predictions about card orders it is these that will be considered here first, accordingly, the observed card orders in these experiments were:

Experiment 4:

p(trop)>p(Euro)>not-q>not-p>q,

Experiment 5 (pooled across rationale and non-rationale conditions):

p>not-q>q>not-p

Experiment 6:

p(trop)=p(Euro)>q>not-q>not-p

Experiment 7:

p>not-q(trop)>not-q(Euro)>q(trop)>q(Euro)>not-p

It can be seen from these figures that generally the observed card orders *do* follow the predictions of Oaksford and Chater's model. However, as with the majority of theories that have been proposed for selection task performance, it is difficult to see how Oaksford and Chater's rational analysis can account for the differences in the selection of the two types of p card (or not-q card in the case of experiment 7). Oaksford and Chater (1994) claim

"There are two dimensions on which the pattern of cards selected in the thematic selection task depends. The first is rule type...The second dimension....is the *perspective* a subject must adopt."

There is no problem with either of these claims as they stand. Rule type clearly has an effect on reasoning as discussed in Chapter 3. The second claim is perfectly reasonable also. Manktelow and Over (1991) clearly demonstrated an effect of perspective on the cards selected by the participants when presented with their version of the selection task (see also Gigerenzer and Hug, 1992; Politzer and Nguyen-Xuan, 1992). However, experiments four to seven reported here would suggest that there is also an effect of probability in the choices which participants make on the deontic selection task. i.e. How *likely* is it that this card will represent a violation of the rule which the participants have been told to check?

Oaksford and Chater do consider probabilistic information in their discussion of selection task literature as they offer a detailed account of the work of Kirby (1994) in which Kirby manipulated the probability of fictional outcomes on the cards with which participants were presented. Kirby used the drinking age rule encountered here earlier. He suggested that the reason the not-q card selections on this version of the task are so high is due to the fact that the probability of a 19 year old illegally drinking alcohol (where an over 21 age limit is present) is also high. He thus added two further not-q cards a 12 year old, and a 4 year old. Kirby predicted that the selections of the not q card would decrease as the age on them decreased. Kirby found support for these predictions. Oaksford and Chater explain this in terms of their model via increased values for the higher ages (M_I = .4 for four year olds, M_I .5 for 12 year olds, and M_I = .6 for 19 year olds), as the 19 and 12 year olds were seen as representing a greater likelihood of violating the rule hence the greater values ascribed to them. Where P(M_I) is the probability of the independence model.

There is no way in the version of the task used here that it is possible to ascribe values such as this using the rule the participants are given. In Kirby's experiment in can be argued that a 19

year old is more likely to be violating the rule. However, it is not the case that someone from a Tropical country is more likely to be violating the rule that '*If you are Entering then you must have cholera among the list of vaccinations.*' The participants were told that cholera is particularly prevalent in Tropical countries, but this does not mean that the passenger from the tropics is any more likely not to be vaccinated against cholera, which would represent a violation of the rule, than a European passenger would. If anything, it could be argued that the Tropical passenger is *more* likely to be vaccinated against cholera than the European passenger simply because of the increased prevalence of the disease in the tropics, and thus less likely to be a violator than the European passenger. It is perhaps true that if the Tropical passenger has not been vaccinated then they represent a greater threat than a European passenger, but this is not the task that the participants have been asked to do. Thus it is difficult to assess how Oaksford and Chater's model could include the necessary probability information in its current formulation. The issue here is that Oaksford and Chater's account would need to include some formulation of perceived risk, which it currently lacks.

In addition to this, Sutherland (1998) has argued that the high levels of not-p card selections on the task are impossible to reconcile within Oaksford and Chater's framework. Table 7.1 shows selections of the Tropical p card as high as 58%, and other values of this card in experiments 8-10 include 44 %, 35%, and 31%. The model provided by Oaksford and Chater gives this card very little value in terms of its ability to reduce uncertainty, and hence increase its likelihood of being selected. Indeed the not-p card is only ascribed a value because all cards have some chance of being selected, hence it is given a value of .1 in Oaksford and Chater's model. It does not seem unreasonable for participants to consider the Tropical not-p card to be of importance given the rationale emphasisng the importance of keeping cholera out of the country.

In summary, the optimal data selection approach is impressive in its ability to account for many of the manipulations of the selection task (Oaksford and Chater, 1994, 1996). However, there remain a number of problems with the theory. It is unable to account for the high level of not-p card selections observed in experiments 8-10. It is also unclear as to how the model would represent the probabilities of the p cards in the deontic experiments here. Finally, as with the discussion of ODS within the indicative framework here, it is difficult to justify a theory of reasoning using Bayes' theorem as a basis, as humans are notoriously poor at applying the theory (Kahneman, Slovic and Tversky, 1982).

8.15 Mental Models

The final theoretical approach to be considered is that of mental models theory (Johnson-Laird, 1983; Johnson-Laird and Byrne, 1991; Johnson-Laird, 1995). The mental models approach is the account favoured here, as it lends itself to the inclusion of other factors, as Manktelow and Over (1991, 1992) have suggested (utility), although the need to revise the theory to include such factors has been disputed by Johnson-Laird and Byrne (1992). As discussed earlier the mental models theory claims that any version of the selection task which leads to an increase in the fleshing out of models will lead to increased performance on the task. The theory would argue that a version such as that used here, based as it is on the immigration task, would lead to increased fleshing out of the models, because it is a deontic task (see Johnson-Laird and Byrne, 1991 p. 81). Participants in the versions reported in experiments 4-10 *did* show increased performance (in comparison with the abstract task) as the theory would predict. The theory would argue that this improved performance was as a result

of a full fleshing out process, leading to the not-q card being explicitly represented and hence more likely to be selected.

As with the other theories considered thus far the issue is how does mental models theory fit with the data from experiments 4 to 10? An important point to note here is that there appears to be two quite distinct patterns observed in the data, as noted by Manktelow, Sutherland and Over (1995). Firstly the participant may select all the Entering cards, regardless of country of origin, or they select only the Tropical Entering cards and none of the European ones, there is very little crossover between these 'two types' of participant. Thus there are two possibilities, either participants see the Tropical cards as more important, and they flesh these models out first, or they see the Entering cards as all the same, and they flesh out a single model for all passengers. It is suggested here that this is the distinction that leads to the pattern of selections observed. What are the models necessary for participants to select the p and not-q cards? How can mental models account for the suppression effect of European p cards?

As was discussed in chapter 2 the model theory suggests that participants construct models of the world or states of affairs that are consistent with the information they have been given in the premises and any relevant knowledge they possess. They then draw a putative conclusion; and proceed to search for counterexamples that are consistent with the premises, but not with the putative conclusion, if none is found then the inference is accepted as valid. Legrenzi, Girotto and Johnson-Laird (1993) have also discussed the issue of focussing in reasoning, and this concept may be of relevance here, i.e. that participants focus their attention on those models that they believe pose the greatest likelihood of violation of the rule. It is important to note here that Evans (1989, 1995) Kirby, (1994) and Sperber, Cara and Girotto (1995) term

this relevance, but as Evans (1995) notes the difference between the two concepts is largely terminological.

When presented with the deontic version of the task Johnson-Laird and Byrne (1995) have argued that participants make their implicit models explicit, and hence correctly solve the task. Johnson-Laird and Byrne argue that participants interpret deontic conditionals as biconditionals. Thus, when presented with the deontic conditional used in experiments 4-10 participants should produce the following models:

e c

¬e ¬c

where e represents ENTERING and c represents cholera. Participants representing the appropriate counterexamples solve the task correctly. For participants in the role of the immigration officer the initial model represented would be:

e c

...

Therefore, the rule (entering and cholera occurring together) is represented explicitly, and the other possible model remains implicit, as shown by the mental footnote '...'. Given this the counterexample represented would be:

e ¬c

This should be the standard selection pattern observed (entering and not cholera), and this is the case, looking at experiment five here, and at Cheng and Holyoak's (1985) results. However, in the experiments reported here there are occasions when the p card is not always selected

(the European p card suppression effect). Is it possible for mental models to account for this suppression effect?

As mentioned earlier, Manktelow, Sutherland and Over (1995) pointed out that there was a clear distinction between two 'types' of participant, those who selected all the p cards, and those who selected only the Tropical p cards. Cummins, Lubart, Alksnis and Rist (1991) presented convincing data regarding disabling conditions with conditional statements. A disabling condition, as defined by Cummins et al. was a contingency that would tend to prevent the inference of the consequent from the antecedent of one of these conditionals. Thus a valid inference is suppressed due to some additional information (this may be supplied via the experimenter as in Byrne, 1989, or rely on the participants' own knowledge). It appears here that the information that a passenger is from a European country provides a disabling condition and thus blocks the fleshing out of the European model, and hence the European entering selections are suppressed. It is possible that the participants who selected all the entering cards merely fleshed out their models without a distinction between Tropical and European. However, those participants who showed suppressed European selections did see coming from Europe as a disabling condition, and hence only fleshed out a Tropical model. This explanation can be supported via evidence from experiment 10 where those participants who were presented with information that showed all rule violators were European had the disabling condition disabled, and hence produced a lessened suppression effect. This suggests an 'integration' of Tropical and European p cards, with cards being treated as entering cards, rather than two distinct types of entering cards.

Johnson-Laird, Legrenzi, Girotto, Sonino-Legrenzi and Caverni (in press) have discussed the role of probability in mental models. They approach this from the perspective that the majority

of participants have little knowledge of probability theory, hence the mistakes typically made on probabilistic tasks (e.g. Kahneman and Tversky, 1983). They refer to such reasoners as naive. A relevant point from this work is that participants construct their mental models, and consider them to be equiprobable by default. However, when given information to the contrary they can adjust the probability values assigned to models.

When the suppression effect is observed in participants, they are given no definitive probability of the likelihood of cholera in Tropical countries, merely told that cholera is particularly prevalent. However, it would seem that this is enough for some participants not to view both types of passengers as 'equiprobable' to be representing a threat to the rule given, hence the suppression effect observed.

It is argued here then that those participants who displayed the suppression effect are representing the counterexample necessary to solve the task correctly (entering and not cholera), but they do not consider this relevant to the European passengers. It is suggested here that they are constructing a counterexample as follows:

		Passenger type
e	¬c	Tropical
...		

Note here that Manktelow and Over (1995) used additional columns to represent utility within Mental Models theory, and Johnson-Laird et al. (in press) to represent explicit frequencies of events. Such a column is suggested now to represent the type of p card to which the model applies. The participant now has models that would lead to the selection of Tropical p cards, and not the selection of European p cards. The first model would yield the observed pattern of selection of Tropical p cards, and not European p cards (as observed), while the second

model notes that there are other possible models that are not explicitly represented here - the same model but relevant to European passengers. Johnson-Laird et al. (in press) note that such mental footnotes as this are easily forgotten due to working memory limitations, hence the observed European p card suppression effect.

However, there are participants who do not display the European p card suppression effect (see section 6.4 for details). As discussed above there are two possible explanations for this pattern of performance. Firstly these participants do not make a distinction between European and Tropical passengers, hence the counterexample model produced by these participants would be:

		Passenger type
e	¬c	All

This model would then lead to the selection of all p cards, regardless of the passenger's country of origin, and this was a recognised pattern of performance in experiment four. Alternatively these participants could flesh out the implicit model represented in their counterexample models, in this case participants would produce the following counterexample:

		Passenger type
e	¬c	Tropical
...		

The first model here would lead participants to select the Tropical p cards. If the implicit model was fleshed out it should yield the following:

		Passenger type
e	¬c	European

This model would then give rise to the selection of the European p cards, as was observed in 11 participants in experiment four (n = 20). It is argued then that if participants hold on to the concept of equiprobability then they should select equal numbers of p cards, regardless of passenger type (as 11 participants did). If participants abandon the notion of equiprobability and consider the Tropical counterexample as more likely to occur, and thus more likely to violate the rule, then they should show the suppression effect (as 9 participants did in experiment four).

An alternative explanation to disabling conditions can be reached via focussing (Legrenzi, Girotto & Johnson-Laird, 1993) although the effect is the same, and it could be argued that the effect of the disabling condition is to focus attention on the Tropical cards. Through the disabling condition the participants focus on the Tropical cards, or see these as more relevant. Subsequently the Tropical models are the only ones fleshed out, and the European selections are suppressed. Alternatively both types (Tropical and European) of model are fleshed out, but because of focussing the Tropical models are given preference, and working memory limitations (Baddeley, 1990) causes the suppression effect to be observed. Either way the suppression effect would be observed as reported in experiments 4 and 7. Again the European bias in experiment 10 would cause a collapse in the disabling condition and hence reduce focussing.

This explanation via focussing can explain the pattern of performance observed. Those participants that produce the counterexample:

		Passenger type
e	¬c	Tropical
...		

Focussing on the first model, and failure to flesh out the second model (represented as a mental footnote) leads to selection of the Tropical p cards only, and the suppression effect for the European p cards, as observed.

The findings from experiment 10 from the final set of experiments are in keeping with these explanations. There was a reduction in the suppression effect in the group that had checked immigration cards that showed European passengers to be violations of the rule. It is argued here that checking through these cards causes an increase in equiprobability of the models (Tropical or European), because the information from the initial set of cards is contrary to the claims in the rationale that cholera is particularly prevalent in Tropical countries. Subsequently participants in this group should represent the counterexample:

		Passenger type
e	¬c	All

This representation would lead to the decrease in the suppression effect that was observed in this group (see section 7.11). The group that checked cards that showed all Tropical passengers to be violations of the rule had their lack of equiprobability confirmed (that Tropical passengers were more risky) and should have produced the following counterexample:

		Passenger type
e	¬c	Tropical
...		

These participants should continue to demonstrate the suppression effect as a consequence of this model and there was support for this prediction.

Again this explanation is consistent with the notion of focussing. The passengers in the Tropical bias group should focus on the Tropical cards because the rationale directs them to such a model, and the information they receive from checking the initial set of cards confirms this. Hence they should represent the counterexamples as:

		Passenger type
e	¬c	Tropical
...		

This should lead to the suppression effect, which there was support for (although the sign test did question this support somewhat).

In the European bias group however, the information on the cards should cause a decrease in focussing, and hence should lead to the model:

		Passenger type
e	¬c	All

Given this model there should be a decrease in the suppression effect for this group, and this was observed.

In summary then the mental models theory gives the best account of the results presented here, through the inclusion of focussing or the representation of probabilistic information in the models. Both of these explanations are consistent with the accounts offered by Johnson-Laird and his colleagues (Legrenzi, Girotto and Johnson-Laird, 1993; Johnson-Laird, Legrenzi, Girotto, Sonino-Legrenzi and Caverni, in press). Indeed it is plausible to claim that the representation of probability information in the models guides focussing.

8.16 Concluding Comments

The experiments reported here have shown that probability has a role to play in reasoning with both indicative and deontic conditionals. In experiments 1-3 it was demonstrated that participants were aware of the importance of the not-q card when they were trying to assess the likelihood of q given p. One of the main outcomes from these initial experiments was the introduction of the Large Array Selection Task. This is a methodological innovation that affords the selection task many attributes unavailable via the standard four-card version. It allows a far more sensitive measure than the standard four-card task when examining the role of probability in reasoning. It also provides a larger amount of data than the standard version. Finally, it avoids the potential pitfalls of cognitive illusions, such as those suggested by Kahneman and Tversky, 1983.

An explanation of these findings was offered in terms of the mental models theory, and other theories of conditional reasoning were deemed either inadequate or inconclusive. The model theory could show that fleshing out was enough to produce the effect of selecting the not-p card in order to make a sound probabilistic judgement, this was not the case with other theories of indicative reasoning.

Following these indicative experiments a move was made to study the influence of reasoning on deontic conditionals, these experiments clearly demonstrated a role for probability in deontic reasoning. An explanation for these findings was given in terms of mental models and disabling conditions (Cummins et al., 1991). The inclusion of disabling conditions also allowed the disappearance of the suppression effect on the European cards to be explained via 'disabling the disabling condition'.

It is suggested here that these experiments and the interpretation of them mark an important step in the integration of reasoning research and decision making research (see Johnson-Laird and Shafir, 1993 and Doherty and Evans, 1996). The experiments reported herein extend the work of Manktelow and Over (1991, 1992) and their use of utility to consider another construct of decision-making research, that of probability. The two fields of decision-making and reasoning have remained separate and distinct until recently. A major contribution to the integration of these two schools has been the recent concept of two forms of rationality (Evans, Over and Manktelow, 1993; Evans and Over, 1996b, 1997). The basic concept here is a distinction between rationality$_1$, which is behaviour that is likely to yield the achievement of goals, and rationality$_2$ that is the degree to which people behave in accordance with normative theories, such as formal logic or probability theory.

The results presented here are in accordance with this distinction. The participants were not behaving in a rational$_2$ manner when the suppression of the European p cards was observed, but it could be interpreted as rational$_1$ behaviour. Participants do not want to select every card, and hence they make decisions to avoid having to do so. By suppressing selection of the European p cards they are behaving in a rational$_1$ way because they are reacting to the rationale - keeping cholera out of the country, and basing their behaviour on this. A criticism of this distinction however, is that it is largely a descriptive framework, and presents little in the way of a testable theory. However, Evans and Over (1996b) presented a re-interpretation of much of the reasoning literature in terms of this distinction with a reasonable amount of success.

Finally, there are a number of issues that remain unresolved. The effect of probability in examples which do not lend themselves so easily to disabling conditions would be useful, as would further exploration of indicative conditionals and probability, via rules which do

produce facilitation even though they are indicative, such as Wason and Shapiro's (1971) Towns and Transport problem.

In terms of deontic conditionals, the majority of research has focussed on permissions and obligations, and there is some confusion as to the nature of these (see Manktelow and Over, 1990). Clearly there needs to be clarification on this issue. A further suggestion here would be an increase in research into other types deontic statements such as promises and threats. In order to increase the links between reasoning and decision-making consideration of probability in other tasks may also prove to be enlightening. Vallée-Tourangeau and New (1997) considered the role of utility in Wason's (1960) 2-4-6 task, and Sutherland, Lucas and Gale (1998) the role of probability. It is only through research of this nature that an integrated theory of reasoning and decision-making can emerge, and through it a full understanding of human thought.

REFERENCES

Almor, A. & Sloman, S.A. (1996). Is deontic reasoning special? *Psychological Review, 103,* 374-380.

Anderson, J.R. (1995). *Cognitive Psychology and its Implications.* New York: W.H. Freeman and Co.

Audi, R. (1989). *Practical reasoning.* London: Routledge and Kegan Paul.

Baddeley, A.D. & Hitch, G.J. (1974). Working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation,* Vol. 8. London: Academic Press.

Baddeley, A.D. (1986). *Working memory.* Oxford: Oxford University Press.

Ball, L.J., Sutherland, E.J., & Quayle, J.D. (1997). Situating deductive competence within a models framework: Some further considerations. *Cahiers de Psychologie Cognitive-Current Psychology of Cognition, 16 (1),* 52-62.

Berry, D.C. (1983). Metacognitive experience and transfer of logical reasoning. *Quarterly Journal of Experimental Psychology, 35A,* 39-49.

Braine, M.D.S. & O'Brien, D.P. (1991). A theory of If: A lexical entry, reasoning program and pragmatic principles. *Psychological Review, 98,* 182-203.

Braine, M.D.S., Reiser, B.J. & Rumain, B. (1984). Some empirical justification for a theory of natural propositional logic. In G.H. Bower (Ed.), *The psychology of learning and motivation,* Vol. 18. New York: Academic Press.

Byrne, R.M.J. (1989). Suppressing valid inferences with conditionals. *Cognition, 31,* 61-83.

Byrne, R.M.J. (1991). Can valid inferences be suppressed? *Cognition, 39,* 71-78.

Byrne, R.M.J. & Handley, S.J. (1992). Reasoning Strategies. *Irish Journal of Psychology: Trinity 400 Special Issue, 13,* 111-124.

Cheng, P.W. & Holyoak, K.J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology, 17,* 391-416.

Cheng, P.W. & Holyoak, K.J. (1989). On the natural selection of reasoning theories. *Cognition, 33,* 285-313.

Cheng, P.W., Holyoak, K.J., Nisbett, R.E. & Oliver, L.M. (1986). Pragmatic versus syntactic approaches to training deductive reasoning. *Cognitive Psychology, 18,* 293-328.

Chorostowski, J.J. & Griggs, R.A. (1985). The effects of problem content, instructions and verbalisation procedure on Wason's selection task. *Current Psychological Research and Reviews, 4,* 99-107.

Cohen, L.J. (1981). Can human irrationality be experimentally demonstrated? *The Behavioural and Brain Sciences, 4,* 317-370.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition, 31,* 187-276.

Cosmides, L. & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. Barkow, L. Cosmides & J. Tooby (Eds.). *The Adapted mind.* Oxford: Oxford University Press.

Cummins, D.D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory and Cognition, 19,* 274-282.

Doherty, M.E. & Evans, J. St.B.T. (1996). Editorial. *Thinking and Reasoning, 2,* 105-108.

Dominowski, R.L. (1995). Content effects in Wason's selection task. In S.E. Newstead & J.St.B.T. Evans (Eds.), *Perspectives on thinking and reasoning: Essays in honour of Peter Wason.* Hove, UK: Lawrence Erlbaum Associates Ltd.

Evans, J.St.B.T. (1977). Linguistic factors in reasoning. *Quarterly Journal of experimental Psychology, 29,* 621-635.

Evans, J.St.B.T. (1983). Linguistic determinants of bias in conditional reasoning. *Quarterly Journal of Experimental*
Psychology, 35A, 635-644.

Evans, J.St.B.T. (1984). Heuristic and analytic processes in reasoning. *British Journal of Psychology, 64,* 391-397.

Evans, J.St.B.T. (1989). *Bias in human reasoning: Causes and consequences.* Hove, UK: Lawrence Erlbaum Associates.

Evans, J.St.B.T. (1991). Theories of reasoning: The fragmented state of the art. *Theory and Psychology, 1,* 83-105.

Evans, J. St. B.T. (1993). The mental model theory of conditional reasoning: Critical appraisal and revision. *Cognition, 48,* 1-20.

Evans, J. St. B.T. (1995). Relevance and reasoning. In S.E. Newstead & J.St.B.T. Evans (Eds.), *Perspectives on thinking and reasoning* (pp. 147-171). Hove, UK: Erlbaum.

Evans, J. St. B.T. (1996). Deciding before you think: Relevance and reasoning in the selection task. *British Journal of Psychology, 87,* 223-240.

Evans, J.St.B.T. (1998). Matching bias in conditional reasoning: Do we understand it after 25 years?. *Thinking and Reasoning, 4,* 45-82.

Evans, J.St.B.T. & Lynch, J.S. (1973). Matching bias in the selection task. *British Journal of Psychology, 64,* 391-397.

Evans, J.St.B.T., Newstead, S.E. & Byrne, R.M.J. (1993). *Human reasoning: The psychology of deduction.* Hove, UK: Lawrence Erlbaum Associates Ltd.

Evans, J.St.B.T., & Over, D.E. (1996a). Rationality in the selection task: epistemic utility versus uncertainty reduction. *Psychological Review, 103,* 356-363.

Evans, J.St.B.T., & Over, D.E. (1996b). *Rationality and reasoning.* Hove, UK: Psychology Press.

Evans, J.St.B.T., & Over, D.E. (1997) Rationality in reasoning: The problem of deductive competence. *Cahiers de Psychologie Cognitive-Current Psychology of Cognition, 16 (1),* 3-38.

Evans, J.St.B.T., & Over, D.E., & Manktelow, K.I. (1993). Reasoning, decision-making and rationality. *Cognition, 49,* 165-187.

Eysenck, M.W. & Keane, M.T. (1990). *Cognitive psychology: A student's handbook.* Hove, UK: Lawrence Erlbaum Associates Ltd.

Fiedler, K. (1988). The dependence of the conjunction fallacy on subtle linguistic factors. *Psychological Research, 50,* 123-129.

Fetzer, J.H. (1993). The argument for mental models is unsound. *The Behavioral and Brain Sciences, 16.*

Garnham, A. & Oakhill, J. (1994). *Thinking and reasoning.* Oxford: Basil Blackwell.

Geis, M.C. & Zwicky, A.M. (1971). On invited inferences. *Linguistic Inquiry, 2,* 561-566.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond "heuristics and biases". In Stroebe & Hewstone (Eds.), *European Review of Social Psychology,* Vol. 2 (pp.83-115). Chichester: Wiley.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky (1996). *Psychological Review, 103,* 592-596.

Gigerenzer, G. & Hug, K. (1992). Domain specific reasoning: Social contracts, cheating and perspective change. *Cognition, 43,* 127-171.

Green, D.W. (1993). Mental models: Rationality, representation and process. *The Behavioral and Brain Sciences, 16,* 352-353.

Green, D.W. (1995). Externalization, counter-examples, and the abstract selection task. *The Quarterly Journal of Experimental Psychology, 48A,* 424-446.

Green, D.W., Over, D.E. & Pyne, R.A. (1997). Probability and choice in the selection task. *Thinking and Reasoning, 3,* 209-235.

Griggs, R.A. (1989). To 'see' or not to 'see': That is the selection task. *Quarterly Journal of Experimental Psychology, 41A,* 517-530.

Griggs, R.A. & Cox, J.R. (1982). The elusive thematic materials effect in the Wason selection task. *British Journal of Psychology, 73,* 407-420.

Hardman, D.K. (1996). Mental models: The revised theory brings new problems. *Behavioral & Brain Sciences, 19(3),* 542-543.

Henle, M. (1962). On the relation between logic and thinking. *Psychological Review, 69,* 366-378.

Holyoak, K.J. & Cheng, P.W. (1995a). Pragmatic reasoning with a point of view. *Thinking and Reasoning, 1,* 289-313.

Holyoak, K.J. & Cheng, P.W. (1995b). Pragmatic reasoning from multiple points of view: A response. *Thinking and Reasoning, 1,* 373-388.

Howson, C. (1997). *Logic with trees.* London: Routledge.

Jackson, S.L. & Griggs, R.A. (1990). The elusive pragmatic reasoning schemas effect. *Quarterly Journal of Experimental Psychology, 42A,* 353-374.

Johnson-Laird, P.N. (1983). *Mental models.* Cambridge: Cambridge University Press.

Johnson-Laird, P.N. (1988). *The computer and the mind: An introduction to cognitive science.* London: Fontana.

Johnson-Laird, P.N. (1995). Inference and mental models. In S.E. Newstead & J.St.B.T. Evans (Eds.), *Perspectives on thinking and reasoning: Essays in honour of Peter Wason.* Hove, UK: Lawrence Erlbaum Associates Ltd.

Johnson-Laird, P.N. & Byrne, R.M.J. (1991). *Deduction.* Hove, UK: Lawrence Erlbaum Associates Ltd.

Johnson-Laird, P.N., & Byrne, R.M.J. (1992). Modal reasoning, models, and Manktelow and Over. *Cognition, 43,* 173-182.

Johnson-Laird, P.N. & Byrne, R.M.J. (1993). Authors' response: Mental models of formal rules? *The Behavioral and Brain Sciences, 16,* 368-376.

Johnson-Laird, P.N. & Byrne, R.M.J. (1995). A model point of view. *Thinking and Reasoning, 1,* 339-350.

Johnson-Laird, P.N., Byrne, R.M.J. & Schaeken, W. (1992). Propositional reasoning by model. *Psychological Review, 99,* 418-439.

Johnson-Laird, P.N., Legrenzi, P., Girotto, V., Sonino-Legrenzi, M., & Caverni, J-P. (in press). Naïve probability: A mental model theory of extensional reasoning. *Psychological Review.*

Johnson-Laird, P.N., Legrenzi, P. & Legrenzi, M.S. (1972). Reasoning and a sense of reality. *British Journal of Psychology, 63,* 395-400.

Johnson-Laird, P.N., & Shafir, E. (Eds.) (1994). Reasoning and Decision Making. Oxford: Blackwell.

Johnson-Laird, P.N. & Taggart, J. (1969). How implication is understood. *American Journal of Psychology, 2,* 367-373.

Johnson-Laird, P.N. & Wason, P.C. (1970). A theoretical analysis of insight into a reasoning task. *Cognitive Psychology, 1,* 134-148.

Kahneman, D., Slovic, P. & Tversky, A. (Eds.) (1982). Judgement under uncertainty: Heuristics and biases. Cambridge: Cambridge University Press.

Kahneman, D. & Tversky, A. (1983). Prospect Theory: An analysis of decision under risk. *Econometrica, 47,* 263-291.

Kahneman, D. & Tversky, A. (1996). On the reality of cognitive illusions : A reply to Gigerenzer's critique. *Psychological Review, 103,* 582-591.

Kirby, K.N. (1994). Probabilities and utilities of fictional outcomes in Wason's four-card selection task. *Cognition, 51,* 1-28.

Kodroff, J.K. & Roberge, J.J. (1975). Developmental Analysis of the conditional reasoning abilities of primary grade children. *Developmental Psychology, 13,* 342-353.

Kosslyn, S.M. (1994). *Image and brain: The resolution of the imagery debate.* Cambridge, MA: MIT Press.

Laming, D. (1996). On the analysis or irrational data selection: A critique of Oaksford and Chater (1994). *Psychological Review, 103,* 364-373.

Legrenzi, P., Girotto, V., & Johnson-Laird, P.N. (1993). Focussing in reasoning and decision making. *Cognition, 49,* 37-66.

Levi, I. (1984) *Decisions and revisions.* Cambridge: Cambridge University Press.

Lowe, E.J. (1993). Rationality, deduction and mental models. In K.I. Manktelow & D.E. Over (Eds.), *Rationality: Psychological and philosophical perspectives.* London: Routledge and Kegan Paul.

Manktelow, K.I. & Evans, J.St.B.T. (1979). Facilitation of reasoning by realism: effect or non-effect?. *British Journal of Psychology, 71,* 227-231.

Manktelow, K.I. & Over, D.E. (1987). Reasoning and rationality. *Mind and Language, 2,* 199-219.

Manktelow, K.I. & Over, D.E. (1990a). Deontic thought and the selection task. In K.J. Gilhooly, M.T.G. Keane, R.H. Logie & G. Erdos (Eds.). *Lines of thinking: Reflections on the psychology of thought, Vol. 1.* Chichester, UK: Wiley.

Manktelow, K.I. & Over, D.E. (1990b). *Inference and understanding.* London: Routledge and Kegan Paul.

Manktelow, K.I. & Over, D.E. (1991). Social roles and utilities in reasoning with deontic conditionals. *Cognition, 39,* 85-105.

Manktelow, K.I. & Over, D.E. (1992). Utility and deontic reasoning: Some comments on Johnson-Laird and Byrne. *Cognition, 43,* 183-186.

Manktelow, K.I., Sutherland, E.J. & Over, D.E. (1995) Probabilistic factors in deontic reasoning. *Thinking and Reasoning, 1,* 201-219.

Markovits, H. (1988). Conditional reasoning, representation, empirical evidence on a concrete task. *Quarterly Journal of Experimental Psychology, 40A,* 483-495.

Marr, D. (1982). *A computational into the human representation and processing of visual information.* San Francisco: W.H. Freeman.

Newstead, S.E., Pollard, P. Evans, J. St.B.T. & Allen, J.L. (1992). The sources of belief bias in syllogistic reasoning. *Cognition, 45,* 257-284.

Oaksford, M.R. & Chater, N. (1993). Reasoning theories and bounded rationality. In K.I. Manktelow & D.E. Over (Eds.). *Rationality.* London: Routledge and Kegan Paul.

Oaksford, M.R. & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review, 101,* 608-631.

Oaksford, M.R. & Chater, N. (1995). Two and three stage models of deontic reasoning. *Thinking and Reasoning, 4,* 350-356.

Oaksford, M.R. & Chater, N. (1996). Rational explanation of the selection task. *Psychological Review, 103,* 381-391.

Oaksford, M., Chater, N., Grainger, B., & Larkin, J. (1997). Optimal data selection in the reduced array selection task (RAST). *Journal of Experimental Psychology: Learning, Memory and Cognition, 23,* 441-458.

Oaksford, M.R. & Stenning, K. (1992). Reasoning with conditionals containing negated constituents. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18,* 835-854.

Oakhill, J. & Johnson-Laird, P.N. (1985). The effect of belief bias on the spontaneous production of syllogistic conclusions. *Quarterly Journal of Psychology, 37A,* 553-570.

O'Brien, D.P. (1993). Mental logic and irrationality: we can put a man on the moon, so why can't we solve these logical reasoning problems? In K.I. Manktelow & D.E. Over (Eds.), *Rationality: Psychological and philosophical perspectives.* London: Routledge and Kegan Paul.

O'Brien, D.P. (1995). Finding logic in human reasoning requires looking in the right places. In S.E. Newstead & J.St.B.T. Evans (Eds.), *Perspectives on thinking and reasoning: Essays in honour of Peter Wason.* Hove, UK: Lawrence Erlbaum Associates Ltd.

Over, D.E. & Manktelow, K.I. (1993). Rationality, utility and deontic reasoning. In K.I. Manktelow & D.E. Over (Eds.). *Rationality.* London: Routledge.

Over, D.E. & Manktelow, K.I. (1995). Perspectives, preferences and probabilities. *Thinking and Reasoning, 1,* 364-371.

Piaget, J. & Inhelder, B. (1969). Intellectual operations and their development. In P. Fraisse & J. Piaget (Eds.). *Experimental Psychology: Its Scope and Method.* London: Routledge and Kegan Paul.

Platt, R.D. & Griggs, R.A. (1993). Facilitation in the abstract selection task: The effect of attentional and instructional factors. *Quarterly Journal of Experimental Psychology, 46A,* 591-614.

Politzer, G. & Nguyen-Xuan, A. (1992). Reasoning about conditional promises and warnings: Darwinian algorithms, mental models, relevance judgements or pragmatic schemas? Quarterly Journal of Experimental Psychology, 44, 401-412.

Polk, T.A. (1993). Mental models, more or less. *The Behavioral and Brain Sciences, 16,* 362-363.

Polk, T.A. & Newell, A. (1995). Deduction as verbal reasoning. *Psychological Review, 102,* 533-566.

Pollard, P. & Evans, J. St. B.T. (1987). On the relationship between content and context effects in reasoning. *American Journal of Psychology, 100,* 41-60.

Popper, K. (1959). *The Logic of Scientific Discovery.* London: Hutchinson.

Pylyshyn, Z.W. (1973). What the mind's eye tells the mind's brain. *Psychological Bulletin, 80,* 1-24.

Quayle, J.D., & Ball, L.J. (1997). Subjective confidence as a determinant of belief bias in syllogistic reasoning. In M.G. Shafto and P. Langley (Eds.), *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society* (pp. 626-631). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Rips, L.J. (1983). Cognitive processes in propositional reasoning. *Psychological Review, 90,* 38-71.

Rips, L.J. (1986). Mental Muddles. In M. Brand & R.M. Harnish (Eds.), *The representation of knowledge and belief.* Tucson: University of Arizona Press.

Rips, L.J. (1990). Reasoning. *Annual Review of Psychology, 41,* 321-353.

Rips, L.J. (1994). *The psychology of proof.* Cambridge, MA: MIT Press.

Rips, L.J. & Conrad, F.J. (1983). Individual differences in deduction. *Cognition and Brain Theory, 6,* 259-289.

Rumain, B., Connell, J. & Braine, M.D.S. (1983). Conversational comprehension processes are responsible for reasoning fallacies in children as well as adults. *Developmental Psychology, 19,* 471-481.

Rumelhart, D.E. (1980). Schemata: The building blocks of cognition. In R.J. Spiro, B.C. Bruce & W.F. Brewer (Eds.). *Theoretical issues in reading comprehension.* Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Snodgrass, J.G. (1977). *The numbers game: Statistics for psychology.* Oxford: Oxford University Press.

Sperber, D., Cara, F. & Girotto, V. (1995). Relevance theory explains the selection task. *Cognition, 57,* 31-95.

Stevenson, R. J. (1993). *Language, thought and representation.* Chichester, UK: John Wiley and Sons.

Stevenson, R. J. & Over, D.E. (1995). Deduction from uncertain premises. *Quarterly Journal of Experimental Psychology, 48A,* 613-643.

Sutherland, E.J. (July, 1998). *Indicative conditionals: Can we account for patterns of reasoning with Bayes' theorem?* Paper presented at the British Psychological Society Post-graduate Affairs Group Conference, University of Derby.

Sutherland, E.J., Lucas, E.J. & Gale, M (September, 1998). *The role of probability in the 2-4-6 task.* Paper presented at the XV British Psychological Society Cognitive Section Conference, University of Bristol.

Tversky, A. & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgement. *Psychological Review, 90,* 293-315.

Vallée-Tourangeau, F. & New, M. (September, 1997). *Optimal hypothesis testing strategies in a rule discovery task.* Paper presented at the XIV British Psychological Society Cognitive Section Conference, University of Bristol, UK.

Von Winterfeldt, D. & Edwards, W. (1986). *Decision analysis and behavioural research.* Cambridge: Cambridge University Press.

Wason, P.C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology, 12,* 129-140.

Wason, P.C. (1966). Reasoning. In B.M. Foss (Ed.), *New horizons in psychology I.* Harmondsworth: Penguin.

Wason, P.C. (1968). On the failure to eliminate hypotheses...A second look. In P.C. Wason & P.N. Johnson-Laird (Eds.), *Thinking and reasoning.* Harmondsworth: Penguin.

Wason, P.C. & Golding, E. (1974). The language of inconsistency. *British Journal of Psychology, 65,* 537-546.

Wason, P.C. & Green, D. (1984). Reasoning and mental representation. *Quarterly Journal of Experimental Psychology, 36A,* 597-610.

Wason, P.C. & Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental; Psychology, 23,* 63-71.

Wildman, T.M. & Fletcher, H.J. (1977). Developmental decreases and increases in solutions of conditional syllogism problems. *Developmental Psychology, 13,* 630-636.

Wilkins, M.C. (1928). The effect of changed material on the ability to do formal syllogistic reasoning. *Archives of Psychology, No. 102.*

Yachanin, S.A. & Tweney, R.D. (1982). The effect of thematic content on cognitive strategies in the four-card selection task. *Bulletin of the Psychonomic Society, 19,* 87-90.

APPENDICES

Appendix 1

Response Sheet Used in Experiment 1

Your father has given you a large amount of money and has asked you to place a bet on a horse for him. He has left it up to you to decide which horse in which race you will bet on.

You are not sure which horse to back however, and your knowledge of horse racing is far more limited than you had led your father to believe. Therefore you do not wish to lose your father's money, because he will be greatly annoyed that you have let him down, and you yourself will look extremely foolish for having lied to your father in the first place. Fortunately a friend tells you about the tipster in "Racing Weekly". She states that **If Racing Weekly tips a horse then it wins.**

Obviously you want to know how confident you can be in your friend's statement, so that you know whether to risk your father's money, and your pride, on Racing Weekly's next tip. Your friend has given you four slips on which she has recorded details of horses she backed in four different races. Each slip says whether the horse was tipped by Racing Weekly or not on one side, and on the other side whether the horse won or not.

Which of the slips would you need to turn over to judge whether your friend's statement will be likely to lead to you not losing your father's bet? Please tick the appropriate answer box(es).

TIPPED BY RACING WEEKLY	NOT TIPPED BY RACING WEEKLY	WON	DID NOT WIN
☐	☐	☐	☐

We don't need to know your personal identity, but the following information will be useful:

Your age..... Sex M..... F..... (please tick)

Have you ever seen this kind of problem before?.....

Thanks.

Appendix 2

Response Sheet Used in Experiment 2

TIPPED BY RACING WEEKLY	NO TIP GIVEN	WON	DID NOT WIN	WAS NOT TIPPED
WAS NOT TIPPED	TIPPED BY RACING WEEKLY	DID NOT WIN	WON	NO TIP GIVEN
NO TIP GIVEN	DID NOT WIN	TIPPED BY RACING WEEKLY	WAS NOT TIPPED	WON
DID NOT WIN	WON	WAS NOT TIPPED	TIPPED BY RACING WEEKLY	NO TIP GIVEN
WON	DID NOT WIN	NO TIP GIVEN	WAS NOT TIPPED	TIPPED BY RACING WEEKLY

Appendix 3

Response Sheet Used in Experiment 3

Week 1	Week 2	Week 3	Week 4	Week 5
WON	LOST	WAS NOT TIPPED	TIPPED BY RACING WEEKLY	LOST
☐	☐	☐	☐	☐

Week 6	Week 7	Week 8	Week 9	Week 10
WON	WAS NOT TIPPED	TIPPED BY RACING WEEKLY	LOST	TIPPED BY RACING WEEKLY
☐	☐	☐	☐	☐

Week 11	Week 12	Week 13	Week 14	Week 15
WAS NOT TIPPED	WON	WAS NOT TIPPED	WAS NOT TIPPED	LOST
☐	☐	☐	☐	☐

Week 16	Week 17	Week 18	Week 19	Week 20
WON	TIPPED BY RACING WEEKLY	WON	LOST	WAS NOT TIPPED
☐	☐	☐	☐	☐

Appendix 4

Response Sheet Used in Experiments 4 & 6

STATUS: ENTERING COUNTRY OF ORIGIN: BRAZIL	STATUS: TRANSIT COUNTRY OF ORIGIN: GERMANY	VACCINATED AGAINST: CHOLERA TYPHOID HEPATITIS	VACCINATED AGAINST: TYPHOID HEPATITIS	STATUS: ENTERING COUNTRY OF ORIGIN: THAILAND
☐	☐	☐	☐	☐
VACCINATED AGAINST: CHOLERA POLIO HEPATITIS	STATUS: TRANSIT COUNTRY OF ORIGIN: BRAZIL	STATUS: ENTERING COUNTRY OF ORIGIN: HOLLAND	VACCINATED AGAINST: TYPHOID RUBELLA	STATUS: TRANSIT COUNTRY OF ORIGIN: NORWAY
☐	☐	☐	☐	☐
VACCINATED AGAINST: POLIO HEPATITIS	VACCINATED AGAINST: RUBELLA MALARIA	STATUS: ENTERING COUNTRY OF ORIGIN: THAILAND	VACCINATED AGAINST: CHOLERA TYPHOID RUBELLA	STATUS: ENTERING COUNTRY OF ORIGIN: SWEDEN
☐	☐	☐	☐	☐
STATUS: ENTERING COUNTRY OF ORIGIN: SWEDEN	VACCINATED AGAINST: TYPHOID HEPATITIS	VACCINATED AGAINST: CHOLERA TYPHOID HEPATITIS	STATUS: TRANSIT COUNTRY OF ORIGIN: BRAZIL	VACCINATED AGAINST: MALARIA CHOLERA
☐	☐	☐	☐	☐

We do not need to know your personal identity, but the following information would be useful:

Your Age..... Sex..... Have you seen this type of problem before?.....

Appendix 5

Response Sheet Used in Experiment 5

ENTERING	TRANSIT	ORIGIN: GERMANY VACCINATIONS: CHOLERA HEPATITIS	ORIGIN: BRAZIL VACCINATIONS: MALARIA RUBELLA	ENTERING
ORIGIN: THAILAND VACCINATIONS: CHOLERA HEPATITIS	ENTERING	TRANSIT	ORIGIN: HOLLAND VACCINATIONS: TYPHOID HEPATITIS	TRANSIT
ORIGIN: GERMANY VACCINATIONS: MALARIA TYPHOID	ORIGIN: BRAZIL VACCINATIONS: CHOLERA RUBELLA	ENTERING	TRANSIT	ORIGIN: GERMANY VACCINATIONS: CHOLERA HEPATITIS
TRANSIT	ORIGIN: HOLLAND VACCINATIONS: CHOLERA MALARIA	ORIGIN: THAILAND VACCINATIONS: RUBELLA HEPATITIS	ENTERING	ORIGIN: BRAZIL VACCINATIONS: MALARIA RUBELLA
ENTERING	TRANSIT	ORIGIN: THAILAND VACCINATIONS: CHOLERA HEPATITIS	ORIGIN: HOLLAND VACCINATIONS: TYPHOID HEPATITIS	

We do not need to know your personal identity, but the following information will be useful:

Your age..... Sex..... Have you ever seen this kind of problem before?

Thanks.

Appendix 6

Response Sheet Used in Experiment 7

ENTERING	TRANSIT	CHOLERA TYPHOID HEPATITIS	TYPHOID HEPATITIS	ENTERING
☐	☐	☐	☐	☐
CHOLERA POLIO HEPATITIS	TRANSIT	ENTERING	TYPHOID RUBELLA	TRANSIT
☐	☐	☐	☐	☐
POLIO HEPATITIS	RUBELLA MALARIA	ENTERING	CHOLERA TYPHOID RUBELLA	ENTERING
☐	☐	☐	☐	☐
ENTERING	TYPHOID HEPATITIS	CHOLERA TYPHOID HEPATITIS	TRANSIT	MALARIA CHOLERA
☐	☐	☐	☐	☐

We do not need to know your personal identity, but the following information will be useful:

Your age..... Sex..... Have you ever seen this kind of problem before?.....
Thanks.

Appendix 7

Screens Used in Experiment 8

VACCINATED: CHOLERA, TYPHOID HEPATITIS	STATUS: TRANSIT ORIGIN: NORWAY	STATUS: ENTERING ORIGIN: HOLLAND	VACCINATED: HEPATTITIS TYPHOID, CHOLERA
VACCINATED: HEPATITIS RUBELLA	STATUS: ENTERING ORIGIN: THAILAND	VACCINATED: TYPHOID RUBELLA	STATUS: TRANSIT ORIGIN: THAILAND
STATUS: TRANSIT ORIGIN: HOLLAND	VACCINATED: CHOLERA HEPATITIS	STATUS: ENTERING ORIGIN: BRAZIL	VACCINATED: RUBELLA CHOLERA
VACCINATED: POLIO, CHOLERA TYPHOID	STATUS: TRANSIT ORIGIN: BRAZIL	VACCINATED: TYPHOID CHOLERA	STATUS: ENTERING ORIGIN: NORWAY

Return to Instructions

Next Page

VACCINATED: TYPHOID HEPATITIS	STATUS: TRANSIT ORIGIN: THAILAND	VACCINATED: RUBELLA TYPHOID	STATUS: TRANSIT ORIGIN: HOLLAND
VACCINATED: TYPHOID RUBELLA	STATUS: ENTERING ORIGIN: BRAZIL	VACCINATED: CHOLERA HEPATITIS	STATUS: ENTERING ORIGIN: THAILAND
STATUS: TRANSIT ORIGIN: THAILAND	STATUS: TRANSIT ORIGIN: NORWAY	VACCINATED: CHOLERA, TYPHOID HEPATITIS	VACCINATED: RUBELLA POLIO
VACCINATED: HEPATTITIS TYPHOID, CHOLERA	STATUS: ENTERING ORIGIN: HOLLAND	VACCINATED: HEPATITIS TYPHOID	STATUS: ENTERING ORIGIN: BRAZIL

Return to Previous Page

Return to Instructions

Finished

Appendix 8

Screens Used in Experiment 9

STATUS: TRANSIT ORIGIN: HOLLAND	VACCINATED: TYPHOID HEPATITIS	STATUS: TRANSIT ORIGIN: BRAZIL	VACCINATED: POLIO HEPATITIS
STATUS: ENTERING ORIGIN: THAILAND	VACCINATED: POLIO HEPATITIS	STATUS: ENTERING ORIGIN: HOLLAND	VACCINATED: HEPATTITIS TYPHOID, CHOLERA
STATUS: TRANSIT ORIGIN: HOLLAND	VACCINATED: CHOLERA RUBELLA	STATUS: ENTERING ORIGIN: THAILAND	VACCINATED: RUBELLA CHOLERA
STATUS: ENTERING ORIGIN: BRAZIL	VACCINATED: CHOLERA HEPATITIS	STATUS: TRANSIT ORIGIN: THAILAND	VACCINATED: HEPATITIS TYPHOID

Next Page

VACCINATED: POLIO, CHOLERA TYPHOID	STATUS: ENTERING ORIGIN: NORWAY	VACCINATED: RUBELLA TYPHOID	STATUS: TRANSIT ORIGIN: THAILAND
VACCINATED: HEPATITIS RUBELLA	STATUS: ENTERING ORIGIN: BRAZIL	VACCINATED: TYPHOID RUBELLA	STATUS: TRANSIT ORIGIN: BRAZIL
VACCINATED: RUBELLA POLIO	STATUS: ENTERING ORIGIN: HOLLAND	VACCINATED: CHOLERA, TYPHOID HEPATITIS	STATUS: TRANSIT ORIGIN: NORWAY
VACCINATED: CHOLERA, TYPHOID HEPATITIS	STATUS: ENTERING ORIGIN: NORWAY	VACCINATED: TYPHOID CHOLERA	STATUS: TRANSIT ORIGIN: NORWAY

Next Page

STATUS: TRANSIT ORIGIN: HOLLAND	VACCINATED: TYPHOID HEPATITIS	STATUS: TRANSIT ORIGIN: BRAZIL	VACCINATED: POLIO HEPATITIS
STATUS: ENTERING ORIGIN: THAILAND	VACCINATED: POLIO HEPATITIS	STATUS: ENTERING ORIGIN: HOLLAND	VACCINATED: HEPATTITIS TYPHOID, CHOLERA
STATUS: TRANSIT ORIGIN: HOLLAND	VACCINATED: CHOLERA RUBELLA	STATUS: ENTERING ORIGIN: THAILAND	VACCINATED: RUBELLA CHOLERA
STATUS: ENTERING ORIGIN: BRAZIL	VACCINATED: CHOLERA HEPATITIS	STATUS: TRANSIT ORIGIN: THAILAND	VACCINATED: HEPATITIS TYPHOID

Next Page

INSTRUCTIONS

Again, please imagine that you are working as an immigration officer at Heathrow airport. You must now check passengers' form H for real, having looked through some sample forms, as you know, one side of the form is the status of the passenger, whether they are Entering the country, or in transit between planes. On the other side of the form is a list of vaccinations that the passengers have had. You are told by the authorities that **If the form says Entering on one side, then Cholera must be among the list of vaccinations on the reverse side.** You are also told that cholera is particularly prevalent in tropical countries.

On the next two pages are 32 such immigration forms, which you must now check for real. In order to ensure that this rule is violated, you must click on those forms that you think you need to check in order to make sure that the above rule is not violated.

You may return to this instruction page at any time, you may now begin the experiment.

Start the experiment

VACCINATED: POLIO, CHOLERA TYPHOID	STATUS: ENTERING ORIGIN: NORWAY	VACCINATED: RUBELLA TYPHOID	VACCINATED: POLIO HEPATITIS
STATUS: ENTERING ORIGIN: THAILAND	VACCINATED: POLIO HEPATITIS	STATUS: ENTERING ORIGIN: HOLLAND	STATUS: TRANSIT ORIGIN: BRAZIL
VACCINATED: RUBELLA POLIO	STATUS: ENTERING ORIGIN: HOLLAND	VACCINATED: CHOLERA, TYPHOID HEPATITIS	STATUS: TRANSIT ORIGIN: NORWAY
STATUS: ENTERING ORIGIN: BRAZIL	VACCINATED: CHOLERA HEPATITIS	STATUS: TRANSIT ORIGIN: THAILAND	VACCINATED: HEPATITIS TYPHOID

Return to Instructions

Next Page

STATUS: TRANSIT ORIGIN: THAILAND	VACCINATED: TYPHOID CHOLERA	STATUS: ENTERING ORIGIN: BRAZIL	STATUS: TRANSIT ORIGIN: HOLLAND
VACCINATED: TYPHOID RUBELLA	STATUS: ENTERING ORIGIN: BRAZIL	VACCINATED: HEPATTITIS TYPHOID, CHOLERA	VACCINATED: HEPATITIS RUBELLA
STATUS: ENTERING ORIGIN: NORWAY	VACCINATED: RUBELLA POLIO	STATUS: ENTERING ORIGIN: THAILAND	VACCINATED: RUBELLA CHOLERA
VACCINATED: CHOLERA RUBELLA	STATUS: ENTERING ORIGIN: NORWAY	VACCINATED: HEPATITIS TYPHOID	STATUS: TRANSIT ORIGIN: BRAZIL

Return to Previous Page

Finished

Appendix 9

Screens Used in Experiment 10

STATUS: ENTERING ORIGIN: HOLLAND	VACCINATED: HEPATITIS CHOLERA, POLIO	VACCINATED: TYPHOID POLIO CHOLERA	STATUS: TRANSIT ORIGIN: NORWAY
VACCINATED: RUBELLA POLIO, CHOLERA	STATUS: ENTERING ORIGIN: THAILAND	VACCINATED: RUBELLA POLIO, TYPHOID	STATUS: TRANSIT ORIGIN: BRAZIL
VACCINATED: HEPATITIS RUBELLA	STATUS: ENTERING ORIGIN: NORWAY	STATUS: TRANSIT ORIGIN: HOLLAND	VACCINATED: TYPHOID, POLIO CHOLERA

Return To Instructions

Next Page

VACCINATED: TYPHOID HEPATITIS	STATUS: TRANSIT ORIGIN: BRAZIL	STATUS: ENTERING ORIGIN: THAILAND	VACCINATED: HEPATTITIS TYPHOID
STATUS: TRANSIT ORIGIN: NORWAY	VACCINATED: CHOLERA HEPATITIS	STATUS: ENTERING ORIGIN: HOLLAND	VACCINATED: POLIO, CHOLERA HEPATITIS
STATUS: TRANSIT ORIGIN: BRAZIL	VACCINATED: POLIO HEPATITIS	VACCINATED: POLIO, HEPATITIS	STATUS: ENTERING ORIGIN: THAILAND

Return To Instructions

Next Page

STATUS: ENTERING ORIGIN: HOLLAND	VACCINATED: HEPATITIS CHOLERA, POLIO	VACCINATED: TYPHOID POLIO CHOLERA	STATUS: TRANSIT ORIGIN: NORWAY
VACCINATED: RUBELLA POLIO, CHOLERA	STATUS: ENTERING ORIGIN: THAILAND	VACCINATED: RUBELLA POLIO, TYPHOID	STATUS: TRANSIT ORIGIN: BRAZIL
VACCINATED: HEPATITIS RUBELLA	STATUS: ENTERING ORIGIN: NORWAY	STATUS: TRANSIT ORIGIN: HOLLAND	VACCINATED: TYPHOID, POLIO CHOLERA

Return To Instructions

Next Page

VACCINATED: TYPHOID HEPATITIS	STATUS: TRANSIT ORIGIN: BRAZIL	STATUS: ENTERING ORIGIN: THAILAND	VACCINATED: HEPATTITIS TYPHOID
STATUS: TRANSIT ORIGIN: NORWAY	VACCINATED: CHOLERA HEPATITIS	STATUS: ENTERING ORIGIN: HOLLAND	VACCINATED: POLIO, CHOLERA HEPATITIS
STATUS: TRANSIT ORIGIN: BRAZIL	VACCINATED: POLIO HEPATITIS	VACCINATED: POLIO, HEPATITIS	STATUS: ENTERING ORIGIN: THAILAND

Return To Instructions

Finished

INSTRUCTIONS

Again, please imagine that you are working as an immigration officer at Heathrow airport. You must now check passengers' form H for real, having looked through all the sample forms, you know, that on one side of the form is the status of the passenger, whether they are Entering the country, or in Transit between planes. On the other side of the form is a list of vaccinations that the passengers have had. You are told by the authorities that **If the form says Entering on one side, then Cholera must be among the list of vaccinations on the reverse side.** You are also told that cholera is particularly prevalent in tropical countries.

On the next two pages are 32 such immigration forms, which you must now check for real. In order to ensure that this rule is not violated, you must click those cards, *and only those cards*, which you think you need to check, in order to make sure that the above rule is not violated. If you make a mistake, then re-click on the card, and it will be de-selected.
You may return to this instruction page at any time.
Thank you, you may now begin the experiment.

Start the experiment

STATUS: TRANSIT ORIGIN: HOLLAND	VACCINATED: RUBELLA, CHOLERA,TYPHOID	VACCINATED: RUBELLA TYPHOID	STATUS: ENTERING ORIGIN: THAILAND
STATUS: ENTERING ORIGIN: HOLLAND	STATUS: TRANSIT ORIGIN: BRAZIL	VACCINATED: TYPHOID RUBELLA	VACCINATED: HEPATITIS TYPHOID, CHOLERA
VACCINATED: RUBELLA POLIO	STATUS: ENTERING ORIGIN: HOLLAND	STATUS: TRANSIT ORIGIN: BRAZIL	STATUS: TRANSIT ORIGIN: NORWAY
STATUS: ENTERING ORIGIN: BRAZIL	STATUS: ENTERING ORIGIN: NORWAY	VACCINATED: TYPHOID CHOLERA	STATUS: ENTERING ORIGIN: THAILAND

Return to Instructions

Next Page

STATUS: ENTERING ORIGIN: THAILAND	VACCINATED: RUBELLA, CHOLERA,TYPHOID	STATUS: ENTERING ORIGIN: HOLLAND	STATUS: TRANSIT ORIGIN: BRAZIL
STATUS: ENTERING ORIGIN: HOLLAND	STATUS: ENTERING ORIGIN: THAILAND	VACCINATED: TYPHOID RUBELLA	VACCINATED: HEPATITIS TYPHOID, CHOLERA
VACCINATED: RUBELLA POLIO	VACCINATED: RUBELLA TYPHOID	STATUS: TRANSIT ORIGIN: BRAZIL	STATUS: TRANSIT ORIGIN: NORWAY
STATUS: ENTERING ORIGIN: BRAZIL	STATUS: ENTERING ORIGIN: NORWAY	VACCINATED: TYPHOID CHOLERA	STATUS: TRANSIT ORIGIN: HOLLAND

Return to Instructions

Finished

PUBLICATIONS

Manktelow, K.I., Sutherland, E.J. & Over, D.E. (1995). Probabilistic factors in deontic reasoning. *Thinking and Reasoning, 1*, 201-219.

THINKING AND REASONING, 1995, 1 (3), 201–219

Probabilistic Factors in Deontic Reasoning

K.I. Manktelow and E.J. Sutherland
University of Wolverhampton, UK

D.E. Over
University of Sunderland, UK

Recent research on reasoning has resulted in a number of authors urging a convergence between ideas in the hitherto disparate fields of deduction and decision making. The deontic reasoning literature in particular refers increasingly to the decision-making constructs of subjective utility and subjective probability. Although the former construct has received some attention from experimenters, the latter has remained relatively unexplored. In this paper a set of experiments is reported in which a modified form of Wason's selection task using an enlarged array was used to investigate the role of a probabilistic factor in reasoning with conditional obligations. Results showed that this factor played a significant role in mediating this reasoning, when probabilistic information was added both to antecedent and to consequent items. Other results indicated that the effect occurred principally by suppressing selections of items with relatively low subjective relevance.

INTRODUCTION

One of the strangest paradoxes of the literature in the psychology of deductive reasoning has been its neglect, until relatively recently, of one of the commonest forms of thought performed by people in everyday contexts: deontic reasoning. Most of the literature has instead been taken up with studies of performance on problems containing indicative sentences, such as standard conditionals of the form *If p then q*. Deontic reasoning can be contrasted with indicative reasoning roughly along the lines of the philosophical distinction between theoretical and practical reasoning: theoretical reasoning aims to infer what is, was, or will be the case; practical reasoning, or deontic reasoning in the present context, aims to infer what one should, may, or must do (Evans, Over, & Manktelow, 1993).

Requests for reprints should be sent to Dr. K.I. Manktelow, School of Health Sciences, University of Wolverhampton, 62–68 Lichfield St., Wolverhampton WV1 1DJ, UK.

The paper was considerably improved by comments from Ruth Byrne, Michael Doherty, and Paolo Legrenzi: grateful thanks to them.

Much of the recent empirical work on reasoning in general has used adaptations of Wason's (1968) selection task. There are differences between the original abstract task and thematic tasks that reliably produce "rational" performance (e.g. D'Andrade's "Sears" version; see Rumelhart, 1980) in these terms: the former requires theoretical reasoning about the existence of a purported state of affairs described by the target conditional, whereas the latter requires practical reasoning about possible actions relative to a regulation whose status is not in question. That such problems as the Sears task are made intelligible by their standing as regulations, rather than by their familiar content (e.g. shops), was established by Cheng and Holyoak (1985).

Cheng and Holyoak's paper was pivotal in many respects: in its findings and their theoretical interpretation, from which most of the recent work on deontic reasoning has taken its cue, and in its methodology, which was adapted for the experiments reported here. In their first experiment, they used two contents, expressing postal and immigration regulations. An example of the latter contained an implied obligation rule (which Cheng and Holyoak termed a permission), *If the form says ENTERING on one side, then the other side includes "cholera" among the list of diseases.* Subjects were cued to the perspective of an immigration officer, checking cards that had either ENTERING (p) or TRANSIT (not-p) on one side, and a list of inoculations either including (q) or not including (not-q) cholera on the other. When a rationale was provided, around 90% of subjects selected the p and not-q cards, the only ones that could contain the critical combination of a passenger entering without a cholera inoculation. Crucially, in a second experiment the detailed content was replaced by reference only to "action A" and "precondition P", in a schematic rule said to be laid down by some authority. There was still a highly significant increase in p, not-q selections compared to a non-deontic condition.

Why and how do deontic contexts give rise to facilitated reasoning performance? This is the issue that is the concern of the present paper. To account for their subjects' behaviour, Cheng and Holyoak introduced the theory of Pragmatic Reasoning Schemas. In brief, this proposed sets of production rules abstracted from experience, which would produce appropriate inferences when a problem contained sufficient cues to activate them. The following were given as the four production rules for regulations like the immigration one:

1. If the action is to be taken, then the precondition must be satisfied.
2. If the action is not to be taken, then the precondition need not be satisfied.
3. If the precondition is satisfied, then the action may be taken.
4. If the precondition is not satisfied, then the action must not be taken.

The target conditional in both experiments matches the structure of rule 1, which is sufficient to cue the schema; rules 1 and 4 are the only ones that

contain determinate statements for action; the p and not-q cards are the only ones that could contain these determinate elements; hence the p and not-q cards are selected.

Pragmatic Reasoning Schema theory was contested first by Cosmides' (1989) social contract theory and later by approaches inspired by the mental models theory of Johnson-Laird and Byrne (1991). Cosmides' approach is derived from evolutionary principles and the idea of social exchange. Parties in a social exchange, she contends, have an understanding that can be expressed in conditional form as *If you take a benefit, then you pay a cost.* She holds that people have an innate tendency to look for those who violate such understandings. These are cheaters, who take benefits without paying the cost. Thus a reasoning task cast in this form should produce a high rate of selections that could detect cheaters. In fact, although there are many results that can be interpreted as attempts to uncover cheaters, there are many that cannot, and the social contract theory has been criticised accordingly (see e.g. Cheng & Holyoak, 1989; Manktelow & Over, 1991, 1995).

However, Cosmides' work was significant for the research reported here in bringing to the fore two factors central to our own approach to deontic reasoning: subjective utility and social cognition. The interplay of these factors was brought out in a set of experiments on deontic selection tasks reported by Manktelow and Over (1991). Predictions about performance in these experiments were made on the basis of a semantic account of deontic conditionals and an associated decision-theoretic analysis of the tasks. Subjects were held to have preferences between mental models of states of affairs from a given social perspective, derived from the problem content and personal knowledge. (See Over & Manktelow, 1993, for more details on this, and Manktelow & Over, 1995, for an even more general analysis.) Construing deontic reasoning in this way has several advantages over the schematic approach: for instance, a semantics based on preferences supplies a reason why a deontic statement should be uttered in the first place. Representation of subjective utility is thus necessary for this form of inference (Manktelow & Over, 1992; Platt & Griggs, 1993).

This account of deontic reasoning relates it closely to the domain of decision making, as we have argued elsewhere (Manktelow & Over, 1991, 1995). Of course, classical decision theory holds that people make decisions on the basis of subjective *expected* utility, i.e. decisions depend on probability judgements as well as ones about utility. Ordinary judgements of both types can be quite vague, and can only rarely be represented by precise numbers, as classical normative decision theory ideally requires. But work on descriptive decision theories, such as that of Kahneman and Tversky (1979), should be directly related to that on ordinary deontic reasoning. Although Manktelow and Over (1991) noted the need to explore probabilistic factors in deontic reasoning, little empirical work has actually been reported so far.

From a more theoretical point of view, Kirby (1994) and Oaksford and Chater (1994) have argued that probability judgements should have an effect in non-deontic selection tasks, based on indicative conditionals. This view is supported by experimental work in Kirby (1994) and by earlier experiments, particularly in Pollard and Evans (1983). To give a full account of these tasks, one must also employ some notion of epistemic utility or of information gain, and so of expected epistemic utility or information gain (Evans & Over, in press; Oaksford & Chater, 1994; Over & Evans, 1994). Thus decision-theoretic ideas are increasingly being applied to understanding selection tasks, both deontic and indicative, although at present this application is controversial. One of the issues this research should help to settle is whether subjects do much indicative or theoretical reasoning as such in selection tasks, or whether they make decisions based purely on expected utility judgements, without performing inferences such as modus ponens or modus tollens. (See Evans, 1995, on this issue).

Probability judgements should indeed affect deductive reasoning performance itself, and not just decision making. This is shown in some recent work on conditional inference. Cummins, Lubart, Alksnis, and Rist (1991) presented causal conditionals whose contents had been rated for numbers of alternative causes of events referred to in the consequents, and number of potential "disabling conditions". A disabling condition was a contingency that, it was predicted, would tend to prevent the inference of the consequent from the antecedent of one of these conditionals. Such inferences were affected by these manipulations. The reason for this, we would suggest, is that the probability of the consequent given the antecedent is affected by the number of disabling conditions and their probability. Whereas Cummins et al. used cues from subjects' general knowledge, Byrne (1989) inserted explicit additional premises into conditional arguments and found that valid inferences could be suppressed. Stevenson and Over (1995) have found a range of suppression effects by varying the degree of uncertainty suggested by additional information, and argue that the suppression occurs because the conditional probability, of the consequent of the conditional given its antecedent, is affected by the extra information.

None of these these studies of inference has concerned deontic reasoning or judgement, and all have used conditional syllogisms rather than the selection task. But probabilistic factors should also affect inferences from, or decisions about, deontic conditionals. In this paper, we report experiments designed to test this possibility. The semantic analysis we have given of deontic thought leads us to predict that people's inferences about probability and utility will interact to determine the selections they make.

EXPERIMENT 1

A probabilistic element was introduced into a deontic selection task about immigration. Subjects were cued to the perspective of an immigration office

checking for potential violators of a conditional obligation specifying required inoculations for entering the UK. It was predicted that the p and not-q cards would be the most frequently selected (representing the possibility of a passenger entering the country but without the requisite inoculation), but that subjects would select more such cards when they indicated that a traveller had come from an area with a higher probability of infection with the specified disease compared with cards for travellers from areas with a lower probability of infection.

Method

Subjects. Twenty first-year undergraduate psychology students from the University of Wolverhampton participated on an unpaid volunteer basis. None had prior experience of the selection task.

Design and Materials. The task materials were adapted from the original immigration content devised by Cheng and Holyoak (1985). The target conditional used in their Experiment 1 was: *If the form says ENTERING on one side then the other side includes cholera among the list of diseases* (see Introduction). This content was changed for a British population and included a rationale (italicised here), as follows:

> Please imagine that you are working as a customs official at Heathrow airport, and as part of your duties you must check the passengers' immigration cards. One side of the immigration card shows whether the passenger is entering Great Britain or if they are just in transit between planes, together with the traveller's country of origin; while the other side of the card shows a list of diseases against which the traveller has been vaccinated.
>
> *You are particularly concerned that people infected with cholera should not be allowed to enter the country. It is well known that cholera is particularly common in tropical countries.*

The rule was *If a person has ENTERING on one side of their immigration card then they must have CHOLERA on the reverse side.* The rationale passage was designed to instil the utility for the detection of potential violations of the rule (first sentence) and to alert the subjects to the greater probability of the disease in some parts of the world (second sentence). Following the task description were instructions for subjects to fill in tick-boxes against the cards they chose (see Appendix).

This experiment differs from other selection tasks previously reported in the use of a enlarged array in which multiple instances of each contingency were presented. Instead of four cards, 20 "immigration cards" were drawn by computer on a single sheet of A4 paper, of which six said ENTERING (p), four

said TRANSIT (not-p), five included (q), and five did not include (not-q) "cholera" among a list of diseases (see Appendix). Using the enlarged array enabled the probabilistic manipulation to be inserted. It consisted of additional information about passengers' country of origin on the p and not-p cards: half gave a tropical country (e.g. Thailand) and half gave a European country (e.g. Germany).

Procedure. The experiment was run during regular classes. Four different orders of cards drawn on the test sheets were used and allotted to subjects at random. Subjects were free to take as long as they liked over the task, and most completed it within 10 minutes. They were asked not to confer, and the experimenter was present to ensure this did not happen. At the end of the test session, the subjects were thanked and given a brief account of the purpose of the experiment.

Results and Discussion

Table 1 shows the percentage frequencies of selections of all cards for all subjects. The numbers of not-p cards were so low that they have been pooled across the probability conditions. It can be seen that there were more overall selections of "Tropical" p cards compared with "European" p cards. The subjects were classified according to the numbers of "Tropical" and "European" p cards each selected: it was predicted that subjects would tend to select more "Tropical" cards. Nine subjects did so, eleven chose equal numbers of each, and none chose more "European" than "Tropical" cards ($P = 0.002$, one-tailed sign test).

Thus the prediction of a difference in selections between p cards on probabilistic grounds has been upheld. It is not possible to compare these results directly with those of Cheng and Holyoak (1985), as these authors used a standard four-card array and did not report single card frequencies. Reading from their graph (1985, p.401) it appears that about 90% of subjects produced

TABLE 1
Experiment 1

p		*not p*	*q*	*not-q*
ENTERING		TRANSIT	"CHOLERA"	"CHOLERA"
Trop.	Euro.	(pooled)	present	absent
98	57	6	5	67

Percentage frequencies of all cards selected by all subjects in Experiment 1. N = 20

For each subject there were 6 p cards available (3 Tropical, 3 European), 4 not-p cards (2 Tropical, 2 European), 5 q cards, and 5 not-q cards.

the p, not-q solution in their immigration task with a rationale. This implies that their subjects produced a higher rate of p, not-q selections overall than did those in the present study. Consequently, it is not possible to tell from the present data whether the probabilistic factor led to a raising of the tendency to select "Tropical" p cards, or a lowering of the tendency to select "European" p cards. It was therefore necessary to run a second experiment including the large array but without the extra information about country of origin on the p and not-p cards.

EXPERIMENT 2

The object of this experiment was to test whether the difference in selection frequencies between the high-probability "Tropical" p cards and the low-probability "European" p cards observed in Experiment 1 reflected a facilitation of selection of the former or a suppression of the latter. The probabilistic information on the p and not-p cards was therefore removed. In addition, the effect of a rationale passage in the instructions, which had been found to affect responses in Cheng and Holyoak's (1985) experiments, was assessed by including it or removing it.

Method

Subjects. Twenty-one first year undergraduate psychology students from the University of Wolverhampton participated on an unpaid volunteer basis. None had prior experience with the selection task.

Design and Materials. The task used the same large array as in Experiment 1, but without the country of origin on the p and not-p cards. The content of the instructional materials was the same, except that for 10 subjects, the rationale passage was included in the instructions, whereas for 11 subjects it was omitted.

Procedure. Subjects were allocated at random to the rationale/no rationale conditions. As before, testing took place during regular classes and was conducted as in Experiment 1.

Results and Discussion

Table 2 gives the percentage selection frequencies for all subjects for the four types of card: each subject could select up to six p cards, four not-p cards, five q cards, and five not-q cards. The main object of interest here is the frequency of selection of the p cards. We can see that they compare closely with the frequency of selection of the "Tropical" p cards in Experiment 1, hence the effect of the probability manipulation in Experiment 1 appears to be to suppress selection of the low-probability p cards rather than to facilitate selection of the high-probability cards. This implies that immigration cards relating to

TABLE 2
Experiment 2

p	*not p*	*q*	*not-q*
ENTERING	TRANSIT	"CHOLERA" present	"CHOLERA" absent
Rationale passage included. N = 11			
80	9	27	53
Rationale passage omitted. N = 10			
98	2	45	50

Percentage frequencies of all cards selected by all subjects in Experiment 2. For each subject in each condition there were 6 p cards available, 4 not-p cards, 5 q cards, and 5 not-q cards.

passengers coming from implicitly low-risk areas were seen as less subjectively significant.

The effects of varying the presence of the rationale passage are not clear. In Cheng and Holyoak (1985), presence of a rationale tended to increase the frequency of p, not-q selections, but there is little evidence of such an effect here, with not-q frequencies around 50% in both conditions, and selections of p cards slightly *lower* with a rationale. There were also more q selections in this experiment than in Experiment 1, indicating that the lack of a clear connection between a rationale and specific probability and utility information produced more variable performance. It is also possible that the enlarged array contributed to this effect, especially compared to standard four-card versions: there were five times as many cards to choose from. The detailed relation between large-array selection tasks and the traditional four-card format will have to await systematic experimental work with indicative sentences and perhaps abstract materials; that was not the focus of the present research.

To explore further the effect of the interaction of the rationale with the information on the cards, a third experiment was run in which the rationale was omitted, but this time the country of origin information was reinstated.

EXPERIMENT 3

The object of this experiment was to test whether the probability effect—the suppression of selection of instances where cholera is less common—requires a link between an explicit rationale setting out the probability and utility structure of the scenario and corresponding information in the problem content. The rationale passage in the instructions, given in the report of Experiment 1

contains both utility and probability elements in its first and second sentences respectively. It states that the subjects are particularly *concerned* to keep people infected with cholera out of their country, and that this disease is particularly *common* in tropical countries. In technical terms, the rationale could be the basis of an expected utility judgement, namely that there is a relatively high expected cost in not examining people from tropical countries. To see whether subjects need to be overtly cued to this rationale, the instruction materials omitted it, and the large array including country of origin information, used in Experiment 1, was employed.

Method

Subjects. Fifteen first-year undergraduate psychology students from the University of Wolverhampton participated on an unpaid volunteer basis. None had prior experience of the selection task.

Design and Materials. The design and materials of Experiment 1 were used, with the exception of the rationale passage in the instructions, which was omitted.

Procedure. Again, the experiment was run during regular classes and conducted as with Experiment 1.

Results and Discussion

Table 3 gives the percentage selection frequencies of all cards for all subjects. With respect to the high/low probability p cards the results could not be clearer: the frequencies were exactly the same, and were at the "non-suppressed" level of the high-probability p cards in Experiment 1. In fact, all but one of the subjects selected all six p cards; the other selected none.

In addition, it can be seen that there is no clear difference between the not-q and the q cards, following the trend established in Experiment 2. It therefore

TABLE 3
Experiment 3

p		*not p*	*q*	*not-q*
ENTERING		TRANSIT	"CHOLERA"	"CHOLERA"
Trop.	Euro.	(pooled)	present	absent
93	93	2	48	33

Percentage frequencies of all cards selected by all subjects in Experiment 3. N = 15.

For each subject there were 6 p cards available (3 Tropical, 3 European), 4 not-p cards (2 Tropical, 2 European), 5 q cards, and 5 not-q cards.

appears once again that explicit representations of utility for at least one of the parties portrayed in a deontic reasoning task are required for clear patterns of inference to be displayed (cf Cheng & Holyoak, 1985; Manktelow & Over, 1992; Platt & Griggs, 1993). In the present case, the same goes for the effects of probability.

One final issue remains for this set of studies: so far, we have reported cases where possibly significant information is on the p and not-p cards. It remains to be seen whether such information on the q and not-q cards will also affect selections.

EXPERIMENT 4

This experiment involved a repeat of Experiment 1, except that the country of origin information was this time given on the q and not-q cards instead of the p and not-p cards. There is some indication from previous work that probability judgements affect inferences that correspond to such choices in the selection task. Byrne (1989) found a marked effect of additional premises on the corresponding modus tollens inference in indicative conditional syllogisms (see also Stevenson & Over, 1995). Strictly speaking, this work is only suggestive. One cannot straightforwardly refer to the modus tollens and modus ponens inferences in an indicative task as equivalent to the corresponding inferences in a deontic task, as deontic conditionals are not truth-functional, as shown by the use of modal auxiliaries in their conclusions. The correspondent to modus tollens here would be to infer "should not enter" from "cholera not mentioned", whereas the correspondent to modus ponens would be to infer "should mention cholera" from "entering". As we also explain later, we doubt that subjects could correctly solve a deontic selection task just by activating some mental deontic logic, which would be a formal way of reasoning taking nothing from the problem content about probability and utility. Subjects could, however, implicitly conform to some of the principles of a deontic logic.

Method

Subjects. Twenty-one first- and second-year undergraduate psychology students from the University of Wolverhampton participated on an unpaid volunteer basis. None had prior experience of the selection task.

Design and Materials. The instructional passages were as in Experiment 1. The card array contained an extra q and not-q card, making six in each case, so that half could be given Tropical and half European countries of origin; there were still six p items and four not-p, but this time without country of origin information.

Procedure. The experiment was again run during regular classes and conducted as with Experiment 1.

Results and Discussion

Table 4 gives the percentage frequencies of selections of all cards for all subjects. The predicted difference between the not-q cards can be seen: there were fewer selections of "European" not-q cards than of "Tropical" not-q cards. Subjects were classified according to whether they chose more "Tropical" or "European" not-q cards, and as predicted, there was a tendency for more subjects to favour the former over the latter: six did so, none did the reverse, and the remainder chose equal numbers ($P = 0.016$, one-tailed sign test). Comparing the not-q frequencies with those in Experiment 1, we see again the apparent difference in subjective relevance between the high-probability "Tropical" cards and the low-probability "European" cards: the latter's selection has been suppressed.

The frequencies of q selections have returned in this experiment to the low levels observed in Experiment 1, compared to the higher levels (around 50%) seen in Experiments 2 and 3. Experiments 1 and 4 are the ones in which all the elements of the task, both the instructional material and the card information, are in place; Experiments 2 and 3 both leave out aspects of one or the other. It would appear, then, that there needs to be clear congruence between the task components for this response, which cannot reveal a potential violation of the stated rule, to be eliminated.

GENERAL DISCUSSION

These experiments clearly demonstrate that probabilistic factors as well as utility judgements play an important role in deontic thought. In Experiment 1 it was shown that the choice of true antecedent cards could be affected by probability and utility information, and Experiment 4 showed that choices of false consequent cards could be similarly affected. Experiments 2 and 3 provided evidence that the effect was to suppress selections of cards about people who

TABLE 4
Experiment 4

p	*not p*	*q*		*not-q*	
ENTERING	TRANSIT	"CHOLERA" present		"CHOLERA" absent	
		Trop.	Euro.	Trop.	Euro.
96	1	16	13	67	40

Percentage frequencies of all cards selected by all subjects in Experiment 4. N = 21.

For each subject there were 6 p cards available, 4 not-p cards, 6 q cards (3 Tropical, 3 European), and 6 not-q cards (3 Tropical, 3 European).

were unlikely to be carrying a disease. The possibility is therefore that selections are being mediated by differing subjective expectations of low-probability and high-probability items: it is not so much that high-probability items somehow acquire extra value, but rather that low-probability items are deemed to be of less value.

One obvious difference between this and other forms of the selection task which needs to be taken into account, as we saw in discussing the results of Experiment 1, is the enlarged array. Its use here was motivated primarily by a methodological concern: to enable us to include a probabilistic variable within the task. However, the selection task normally only involves "one shot" at each contingency. The only exception in the literature is the so-called RAST (Reduced Array Selection Task), in which subjects are presented with repeated instances of just the q and not-q items, hence its name. One property of this form of the task is that most subjects at some stage produce "facilitated" performance, i.e. cease examining the q items and focus just on the not-q items (see Wason & Green, 1984). Note that the RAST asks subjects to make repeated observations concerning an inference, and therefore differs radically from the requirements of the standard selection task. Moreover, Gigerenzer (e.g. 1993) has argued persuasively that people are natural "frequentists" in their decision making: their behaviour corresponds more closely to normative prescriptions when making judgements about frequencies than about single events. These considerations would lead one to expect, all else being equal, "facilitated" performance with large array selection tasks; however, we have seen that subjects' choices are still prone to probabilistic effects. This conclusion is not damaged by the differences between the standard task and the large-array version used here.

The rationale manipulated in these experiments made statements about both utility and probability—it referred to a concern to keep cholera out of the country and claimed that this disease is common in tropical countries. It may be that stressing the utility side of this rationale alone would have had some effect. Subjects might not need to be given frequency information, about where cholera is common, for them to make appropriate probability judgements and to tend not to choose "European" p cards. They might have heard several reports of cholera in tropical countries, or even just read one particularly striking report, and consequently judge that people from such countries are more likely to have cholera than people from European countries. That is, an availability or vividness heuristic might lead subjects to make this judgement even in the absence of frequency information. The effect of these possible heuristics in deontic selection tasks will have to be examined in future work, and could provide a further link with the decision-making literature. (See Kahneman, Slovic, & Tversky, 1982, on these heuristics.)

Our findings concur with the results of the only other study of probabilistic effects in deontic selection tasks known to us. Kirby (1994) used the drinking-

age task devised by Griggs and Cox (1982). Subjects were cued to the perspective of a security officer checking for violations of the following rule: *If a person is drinking beer, then the person must be over 21 years of age.* Originally, the not-q card showed an age near to that specified in the rule, e.g. 19. Kirby ingeniously added two other not-q cards, showing "12 years" and "4 years". He predicted, and observed, that there would be fewer selections of not-q cards as the stated age decreased, because subjects would simply not see the point of checking whether a young child was drinking beer: it is an intrinsically unlikely occurrence, in most people's experience. Here again we have a probability judgement combined with a utility one: the expected cost of missing a 4-year-old beer drinker is less in Kirby's rationale than that of missing a 12- or 19-year-old beer drinker.

Kirby discusses his results in terms of subjective relevance, which has also been applied to selection tasks by Evans (1989, 1995), Legrenzi, Girotto, and Johnson-Laird (1993), and Sperber, Cara, and Girotto (in press). Legrenzi et al. prefer the term "focusing", but Evans (1995) argues that the differences between their idea and that of relevance is largely terminological. For Legrenzi et al., subjects in a deontic selection task focus on, or see as relevant, cards that might reveal a violation of the rule they are given. This means that the subjects explicitly represent the violations in their mental models, and that is why they select the cards that they do. We would ask why the focusing happens in the first place, or why certain possible outcomes are seen as relevant to a question about violation. There has to be something significant in a "violation" before a question about it has the right effect (Kirby, 1994; Manktelow & Over, 1992). Sperber et al. put it in a word by saying that the violation cases are made relevant by being made "undesirable" in some rationale. In more technical but equivalent terms, these cases have relatively low utility. What our experiments here show is that relatively low *expected* utility also has an effect. Even more generally, what is probabilistically relevant, i.e. information that raises or lowers the probability of a proposition in question, is bound to be a major factor in what is pragmatically relevant.

Legrenzi et al. (p. 39) make the point that human reasoners seem to be "inferential satisficers", who will not go on thinking once they find a conclusion which "fits the available facts", with the "potentially disastrous consequence of overlooking the correct conclusion". In the deontic domain we have explored here, the potential error is that of letting a person with a deadly disease, acquired by chance in a low-risk area, enter the country. Of course, one could argue that it is sometimes rational, and an efficient use of one's time, to ignore the possibility of such an error if its utility × its probability, i.e. its expected cost, is low enough.

There is evidence in our data that there are two distinct groups of subjects: those who will tend to turn over just "Tropical" p or q cards, and those who will tend to turn over all the p or q cards. About half the subjects in Experiments 1

and 4 fell into either category. One group of subjects might have interpreted the rule they were given as being without exceptions. The other might have inferred from the context that they were being implicitly given two rules: one stating that they had an obligation to inspect visitors from tropical countries, and the other that they just had permission to check those from European countries. If this is the right way to look at the experiments, we would ask why these different interpretations existed. All the subjects are asked, by the rationale, to take the point of view of a customs officer, and the difference between them could reflect judgements about how such officers should rationally allocate their time. That is, the different groups of subjects might have had different interpretations of the rule because of their different expected utility judgements. It will be interesting to explore the psychological differences between these groups in future research; at present, generalisation from these findings must be treated cautiously in view of the one type of content we have used. An obvious step would be to construct scenarios that were less dependent on subjects' world knowledge, as did Cheng and Holyoak (1985) and Manktelow and Over (1991), among others.

Our current findings are clearly consistent with the views of Legrenzi et al. (1993), and provide evidence for probability judgements as an important influence on which items become focused upon, and hence explicitly represented. In mental model terms, it appears that the relevance of the low-probability items remains implicit for some subjects, while high-probability items form part of an explicit representation that delivers the required selections. We favour an account of deontic thought that employs mental models, at least in part. (See Manktelow & Over, 1995 for more details on our approach, which is influenced by that of Johnson-Laird & Byrne, 1991, but distinct from it, as is clear from Johnson-Laird & Byrne, 1992, and Manktelow & Over, 1992.) Mental models represent states of affairs, and can be naturally extended to represent preferences between those states and the probabilities of them.

To follow this approach, however, one does not have to suppose implausibly that people naturally assign numbers for utilities and probabilities to the states they represent, nor that they always conform to the probability calculus or normative decision theory. People could give some vague order to their mental models without using precise numbers, but indicating that some of these were more preferable or probable than others. They might also tend to focus on costly outcomes that they seem at risk of suffering, in order to avoid them or take corrective action. That would give them some tendency to loss aversion in deontic selection tasks, and there is some evidence that this exists. (See Manktelow & Over, 1990, and Kirby, 1994; on loss aversion in a descriptive decision theory, see Kahneman & Tversky, 1979.)

Some other theoretical points can be made in conclusion. The probabilistic effects demonstrated here, along with those of subjective utility and social role demonstrated in earlier research (reviewed in Manktelow & Over, 1995), pose problems for some of the approaches that have previously been invoked to

explain performance in realistic forms of the selection task. For instance, the results are very hard to interpret from a purely mental-logic point of view. One of the main advocates of this view has argued that the standard selection task has nothing to say about logical reasoning because it is too difficult to enable subjects to parade their skills (O'Brien, 1995). However, he says little about those forms of the task, principally deontic forms, in which people do produce clear selection patterns that can be shown to be rational, to some extent, by a decision-theoretic analysis (see Evans et al., 1993).

It is unclear how a mental logic approach could accommodate the demonstrated effects of subjective probability, as well as those of subjective utility and social perspective. There are indeed deontic logics, but these sanction few inferences, and a mental version of any we know would not go very far in explaining the richness of ordinary deontic reasoning. This kind of practical reasoning is often, we repeat, much more closely related to the kind of judgement found in decision making than it is to what mental logic theorists call "reasoning". It must go way beyond mere formal, content-independent inference, as it can only be done well by making probability and utility judgements about content-rich special cases. (See Stevenson & Over, 1995, on the problems mental logic has in accounting for the way probability judgements affect indicative or theoretical reasoning.)

The approach that inspired much of the research on deontic reasoning, that of Pragmatic Reasoning Schemas, initiated by Cheng and Holyoak (1985), is similarly silent on the effects of probability. However, it does hold that regulations are laid down by authorities to achieve goals, such as that of keeping deadly diseases out of a country. The production rules in the schemas refer to *actions* in their consequents that are to be taken as steps towards these presupposed goals. This aspect of this work has received little attention in the literature, although it has recently become more explicit (Holyoak & Cheng, 1995). It is through the idea of a goal underlying deontic reasoning that an extension of this approach can be related to decision making, and could possibly account for our results. A goal people have in some context is simply an outcome they prefer to others, and one that they think they will probably attain by taking certain steps. It may be that production rules are not activated, in some context, if the steps they refer to seem unlikely to lead to the given goal. Thus some subjects might not have used certain production rules in our experiments because they judged that the goal of keeping cholera out the country was not best served by examining "European" cards.

It is true that, in our scenario, no qualifications are made to the immigration regulation. Even so, authorities who lay down rules often expect people to make qualifications and work out exceptions to the rules for themselves. Without this, many rules would be tremendously complex, and applying them without exception would still be counterproductive at times. The goal of most traffic regulations, for example, is to prevent death and injury on the roads while

keeping vehicles moving. These cannot be too complex if they are to be remembered quickly and applied efficiently, but no matter how complex they became, there would always be cases when breaking them would prevent an accident that conforming to them would cause. In more technical words, good deontic reasoning is highly non-monotonic, and the "disabling conditions" for it are excuses and mitigating circumstances (Over & Manktelow, 1993). General rules can be laid down and applied easily in many circumstances, but sometimes we need to withdraw a deontic conclusion we have inferred from the rules after learning more about a particular case. People do this, in that case, precisely to help them achieve the goals that the rules are there to serve in general. They are able to do this, we would contend, because they can grasp the goals, by modelling states of affairs, and can make special utility and probability judgements to achieve them in exceptional cases.

Whether the type of production rules employed by Cheng and Holyoak can fully account for deontic reasoning is still an open question. Our results here show that this approach has to take account, in some way, of probabilistic factors, and it has to find a way of making its production system non-monotonic. There are also questions about the production rules themselves, which seem to be more like deontic conditionals than rules in a standard production system. Cheng and Holyoak's production rules have the deontic terms "must" and "may" in them, and some of the antecedents begin with "if the action is to be taken . . .", which is an implicitly deontic construction. It is unclear what logical or computational system underlies these rules (Over & Manktelow, in press). Finally, this whole approach seems to imply that subjects examine all the cards and perform deontic inferences about all of them. There is evidence that subjects do not do this, but instead focus on or see as relevant only those cards that may reveal a significant violation (Evans & Clibbens, in press).

People may well make use of some kind of schemas in their deontic reasoning, but our position is that at least the underlying representations are best captured by mental models. We hold that people need to model states of affairs to have preferences among them and to make probability judgements about them. They are also able to model the preferences and probability judgements of other people, with other views or perspectives, including those with authority and power. They then have goals, whether directly their own or taken from authorities, for the actions they perform. Any possible schemas, inferences, or actions in the world are seen as significant if they seem likely to help to achieve these goals. This whole process is at the basis of deontic thought: it tells us what rules to accept, and when exceptions to them are appropriate. Throughout this process, subjective judgements are necessary. That is why people can differ so much in their deontic thought, as revealed in these experiments in the different ways subjects respond to "European" cards. Obtaining a proper account of this process will only be possible if work on deontic reasoning is further integrated with that on decision making.

Manuscript received 9 September 1994
Revised manuscript received 18 April 1995

REFERENCES

Byrne, R.M.J. (1989). Suppressing valid inferences with conditionals. *Cognition, 31* 61–83.

Cheng, P.W., & Holyoak, K.J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology, 17,* 391–416.

Cheng P.W., & Holyoak, K.J. (1989). On the natural selection of reasoning theories. *Cognition, 33,* 285–313.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition, 33,* 187–276.

Cummins, D.D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory and Cognition, 19,* 274–282.

Evans, J.StB.T. (1989). *Bias in human reasoning: Causes and consequences.* Hove, UK: Lawrence Erlbaum Associates Ltd.

Evans, J.StB.T. (1995). Relevance and reasoning. In S.E. Newstead & J.StB.T. Evans (Eds.), *Current directions in thinking and reasoning.* Hove, UK.: Lawrence Erlbaum Associates Ltd.

Evans, J.StB.T., & Clibbens, J. (in press). Perspective shifts in the selection task: Reasoning or relevance? *Thinking and Reasoning.*

Evans, J.StB.T., & Over, D.E., (in press). Rationality in the selection task: Epistemic utility versus uncertainty reduction. *Psychological Review.*

Evans, J.StB.T., Over, D.E., & Manktelow, K.I. (1993). Reasoning, decision making and rationality. *Cognition, 40,* 165–187.

Gigerenzer, G. (1993). The bounded rationality of probabilistic mental models. In K.I. Manktelow & D.E. Over (Eds.), *Rationality.* London: Routledge.

Griggs, R.A., & Cox, J.R. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology, 73,* 407–420.

Holyoak, K.J., & Cheng, P.W. (1995). Pragmatic reasoning about human voluntary action: Evidence from Wason's selection task. In S.E. Newstead & J.StB.T. Evans (Eds.), *Current directions in thinking and reasoning.* Hove, UK: Lawrence Erlbaum Associates Ltd.

Johnson-Laird, P.N., & Byrne, R.M.J. (1991). *Deduction.* Hove, UK: Lawrence Erlbaum Associates Ltd.

Johnson-Laird, P.N. & Byrne, R.M.J. (1992). Modal reasoning, models, and Manktelow and Over. *Cognition, 43,* 173–182.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgement under uncertainty: Heuristics and biases*: Cambridge: Cambridge University Press.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47,* 263-291.

Kirby, K.N. (1994). Probabilities and fictional outcomes in Wason's four-card selection task. *Cognition, 51,* 1–28.

Legrenzi, P., Girotto, V., & Johnson-Laird, P.N. (1993). Focussing in reasoning and decision making. *Cognition, 49,* 37–66.

Manktelow, K.I., & Over, D.E. (1990). Deontic thought and the selection task. In K.J. Gilhooly, M.T. Keane, R.H. Logie, & G. Erdos (Eds.), *Lines of thinking, Vol. 1.* Chichester, UK: Wiley.

Manktelow, K.I., & Over, D.E. (1991). Social roles and utilities in reasoning with deontic conditionals. *Cognition, 39,* 85–105.

Manktelow, K.I., & Over, D.E. (1992). Utility and deontic reasoning: Some comments on Johnson-Laird and Byrne. *Cognition, 43,* 183–188.

Manktelow, K.I., & Over, D.E. (1995). Deontic reasoning. In S.E. Newstead & J.StB.T. Evans (Eds.), *Current directions in thinking and reasoning.* Hove, UK: Lawrence Erlbaum Associates Ltd.

Oaksford, M.R., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review, 101,* 608–631.

O'Brien, D.P. (1995). Finding logic on human reasoning requires looking in the right places: A mental-logic view of logical-reasoning tasks. In S.E. Newstead & J.St.B.T. Evans (Eds.), *Perspectives on thinking and reasoning*. Hove, UK: Lawrence Erlbaum Associates Ltd.

Over, D.E., & Manktelow, K.I. (1993). Rationality, utility and deontic reasoning. In K.I. Manktelow & D.E. Over (Eds.), *Rationality: Psychological and Philosophical Perspectives*. London: Routledge.

Over, D.E., & Manktelow, K.I. (in press). Perspectives, preferences, and probabilities. *Thinking and Reasoning*.

Platt, R.D., & Griggs, R.A. (1993). Darwinian algorithms and the Wason selection task: A factorial analysis of social contract selection problems. *Cognition, 48*, 163–192.

Pollard, P., & Evans, J.StB.T. (1983). The effect of experimentally contrived experience on reasoning performance. *Psychological Research, 45*, 287–301.

Rumelhart, D.E. (1980). Schemata: The building blocks of cognition. In R.J. Spiro, B.C. Bruce, & W.F. Brewer (Eds.), *Theoretical issues in reading comprehension*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Sperber, D., Cara, F., & Girotto, V. (in press). Relevance theory explains the selection task. *Cognition*.

Stevenson, R.J., & Over, D.E. (1995). Deduction from uncertain premises. *Quarterly Journal of Experimental Psychology, 48A*, 613–643.

Wason, P.C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology, 20*, 273–281.

Wason, P.C., & Green, D.W. (1984). Reasoning and mental representation. *Quarterly Journal of Experimental Psychology, 36A*, 597–610.

APPENDIX

The enlarged array of cards used in Experiment 1, with country or origin added to the p and not-p items. In Experiment 4, this information was added instead to the q and not-q items.